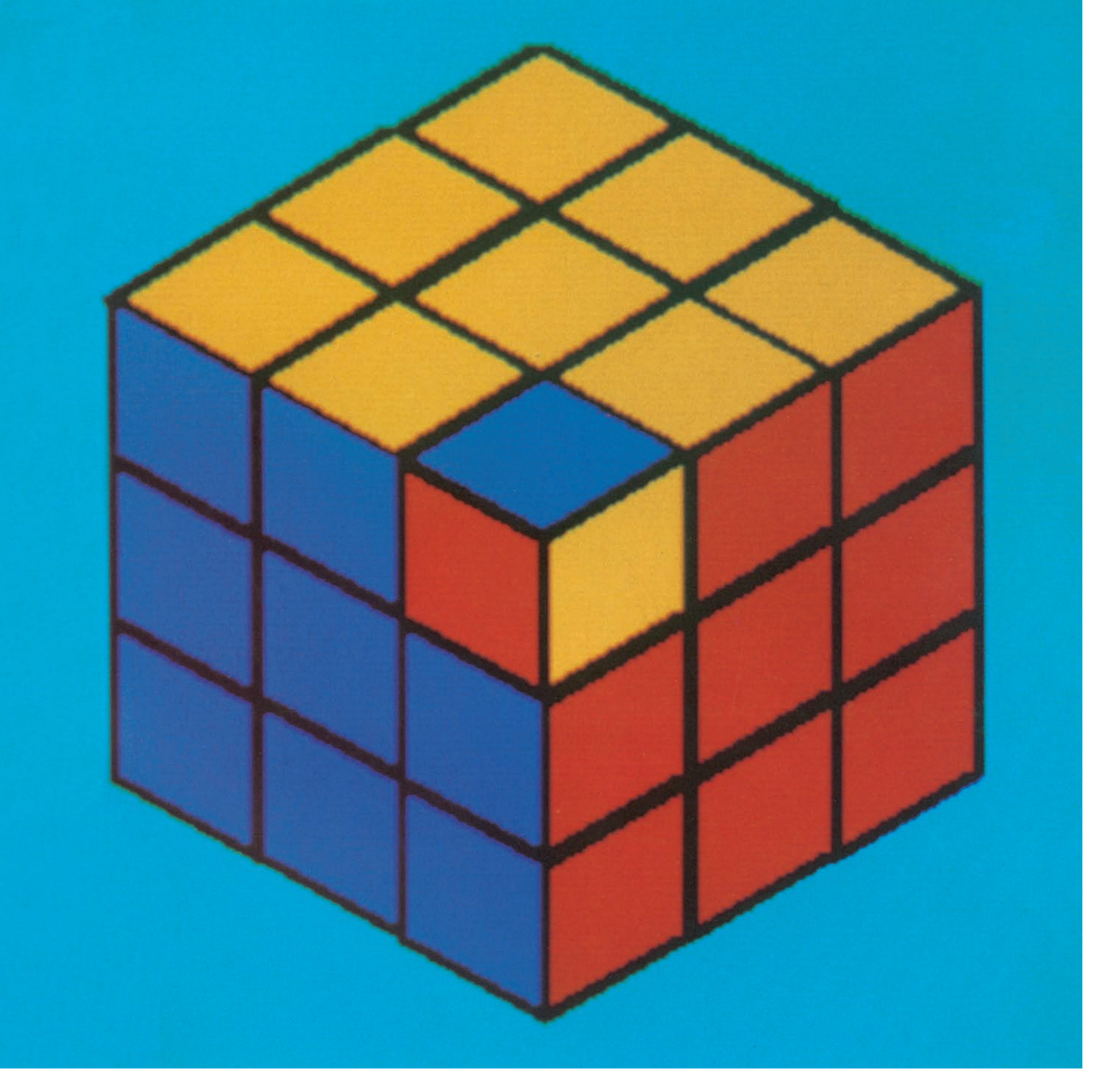
INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN



CUBOLOGIA MAGICA

Mayo 1981

250 PTAS.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 56
Mayo 1981

INVESTIGACION Y
CIENCIA *Edición en español de* SCIENTIFIC AMERICAN

EDITA: Prensa Científica S.A.
Calabria, 235-239
Barcelona-29 (ESPAÑA)

Colaboradores de este número:

Asesoramiento y traducción:

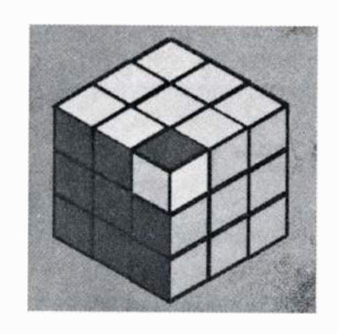
Ramón Fontarnau: *El armazón celular;* Miguel Gich: *Las erupciones del monte Santa Elena;* Manuel Puigcerver: *La galaxia de la Vía Láctea;* G. Nicolás Rodrigo: *Polinización submarina;* J. Manuel García de la Mora: *Percepción anortoscópica;* Julio Rodríguez Villanueva: *Bioquímica de la resistencia a la malaria;* Manuel García Doncel: *El descubrimiento newtoniano de la gravitación;* Luis Bou: *Temas metamágicos;* J. Vilardell: *Taller y laboratorio.*

Ciencia y sociedad:

Diego Gracia, Pascual Cuñat, A. Aguilar y V. García

Libros:

María Rosa Miracle, Salvador Alegret y Jaume Terradas



LA PORTADA

La ilustración de portada, obtenida por ordenador, nos presenta un Cubo Mágico, nuevo rompecabezas que está en nuestros días volviendo locos a un número cada vez mayor de cubo-adictos de todo el mundo (*véase la sección "Temas metamágicos"*). En su estado fundamental (no el de portada) cada cara del cubo está formada por nueve "cubitos" del mismo color. Cada plano de nueve cubitos puede girar alrededor del central; por ello, bastan unos cuantos giros para dejar totalmente entremezclados sus colores. El problema fundamental consiste en devolver al cubo su ordenación inicial. El número de posibles giros del proceso de reordenación es astronómicamente grande y la resolución del problema requiere perspicaz análisis. En la portada vemos un cubito-vértice que ha sufrido la torsión de 120 grados en sentido horario, mientras todos los demás permanecen en su estado inicial. Tal situación, imposible de materializar, ha sido bautizada "quark", en recuerdo de la inaccesible partícula de igual nombre.

Suscripciones:
Prensa Científica, S. A.
Calabria, 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 41

Condiciones de suscripción:
España:
Un año (12 números): 2.750 pesetas
Extranjero:
Un año (12 números): 43 U.S.$
Ejemplar atrasado ordinario:
280 pesetas
Ejemplar atrasado extraordinario:
420 pesetas

Distribución para España
Distribuciones de Enlace, S. A.
Ausias March, 49, Barcelona-10

Distribución para los restantes países:
Editorial Labor, S. A.
Calabria, 235-239 - Barcelona-29

Publicidad:
Madrid:
Gustavo Martínez Ovin
Avda. de Moratalaz, 137, Madrid-30
Tel. 430 84 81
Cataluña:
Miguel Munill
Balmes, 191, 2.°, 2.ª, Barcelona-6
Tels. 218 44 45 y 218 40 86

Controlado por O.J.D.



PROCEDENCIA DE LAS ILUSTRACIONES

Fotografía de la portada de Ralph Morse

Página	Fuente
7	Luis Castañer y Miguel Alonso
8	Manuel Puigcerver y Evangelina Oriol
9-14	Luis Castañer y Miguel Alonso
16-18	Keith R. Porter
19	Tom Prentiss
20-27	Keith R. Porter
31	U.S. Geological Survey
32-33	Andrew Tomko
34	Robert Decker
36-37	National Aeronautics and Space Administration
38-39	Robert Decker
40-44	Andrew Tomko
54-56	Allen Beechel
57	United Kingdom-European Southern Observatory Sky Survey (*arriba*), Allen Beechel (*abajo*)
58-60	Allen Beechel
61	Mount Stromlo Observatory
62	Observatorio de Las Campanas
64	Allen Beechel
65	Observatorio Interamericano de Cerro Tololo
66	Allen Beechel
68-69	John N. Bahcall y Raymond M. Soneira
70	Palomar Sky Survey
74	John Pettitt
76-78	Ilil Arbel
79-80	John Pettitt
87-94	Jerome Kuhl
99	James B. Jensen (*arriba*), William Trager (*abajo*)
100-103	Patricia J. Wynne
104	Milton J. Friedman
105-106	Patricia J. Wynne
110	Jeremy Whitaker
112	Alan D. Iselin
113	Historical Society of Pennsylvania
114	Alan D. Iselin
115	University Library, Cambridge
116	Goldsmith Company's Library, University College, Londres
118-119	Alan D. Iselin
123-133	Alan D. Iselin
139-146	Michael Goodman

ISSN 0210-136X
Dep. legal: B. 38.999-76
Fotocomposición Tecfa
Guipúzcoa, 36 (local 1) Barcelona-20
Fotocromos reproducidos por GINSA, S. A.
Imprime GRAFESA
Gráfica Elzeviriana, S. A.
Nápoles, 249 - Tel. 207 40 11
Barcelona-13
Printed in Spain - Impreso en España

Los autores

LUIS CASTAÑER ("Energía solar fotovoltaica") nació en Zaragoza y cursó estudios en la Universidad Politécnica de Madrid, por la que es ingeniero superior de telecomunicaciones. Fue becario de la Agencia Espacial Europea y realizó trabajos de investigación en calidad de tal en la Universidad Paul Sabatier, de Toulouse, donde obtuvo el doctorado en 1973. Es también doctor, con premio extraordinario, por la Universidad Politécnica de Barcelona. Becario del Comité Conjunto Hispano Norteamericano en la Brown University a lo largo del curso 1979. Desde 1976 es catedrático interino de electrónica de la Escuela Técnica Superior de Ingenieros de Telecomunicaciones de Barcelona, de la que fue subdirector de investigación desde julio de 1977 hasta diciembre de 1978. Dirige y realiza trabajos de investigación y desarrollo en energía solar fotovoltaica.

KEITH R. PORTER y JONATHAN B. TUCKER ("El armazón celular") abordan el tema de su artículo partiendo de unas carreras profesionales muy diferentes. Porter, pionero en la aplicación del microscopio electrónico al estudio de las estructuras celulares, enseña biología celular en la Universidad de Colorado en Boulder. Tras iniciar sus estudios en la Universidad de Acadia, pasó luego a la de Harvard, donde se doctoró. Durante más de veinte años estuvo en el Instituto Rockefeller para la Investigación Médica; allí llegó a ser profesor y jefe del departamento de citología. En 1961 se puso al frente del departamento de biología de la Universidad de Harvard y, en 1974, del departamento de biología molecular, celular y del desarrollo de la de Boulder. Tucker, que se especializó en biología en la Universidad de Yale, trabaja por libre.

ROBERT DECKER y BARBARA DECKER ("Las erupciones del monte Santa Elena") han colaborado en varios libros y artículos sobre los volcanes. Ambos participaron en la investigación, sobre el terreno, de la zona del monte Santa Elena durante la primavera y el verano de 1980. En ese intervalo, Robert Decker ostentó la responsabilidad científica del Proyecto monte Santa Elena, acometido por el Servicio Geológico de los Estados Unidos. Estudió geología y geofísica en el Instituto de Tecnología de Massachusetts y en la Escuela de Minas de Colorado, doctorándose por esta última institución. Fue presidente de la Asociación Internacional de Vulcanología y Química del Interior de la Tierra. Antes de trabajar para el Servicio Geológico, en 1979, enseñó geofísica en el Dartmouth College. Bárbara Decker, licenciada en periodismo por la Universidad de California en Berkeley, es publicista. El matrimonio Decker vive actualmente en las islas Hawai, donde él dirige la Estación de Observación Vulcanológica de las islas Hawai, adscrita al Servicio Geológico de los Estados Unidos.

BART J. BOK ("La galaxia de la Vía Láctea) es profesor jubilado de astronomía de la Universidad de Arizona. Nacido y educado en Holanda, recibió su doctorado en la Universidad de Groningen. Ingresó en la de Harvard en 1929 como beneficiario de una ayuda Robert Wheeler y Willson para astronomía; allí permaneció hasta convertirse en profesor de astronomía de la cátedra Robert Wheeler Willson y subdirector del Observatorio Harvard. Deja estos cargos al trasladarse a la Universidad Nacional Australiana y asumir la dirección del Observatorio del monte Stromlo. En 1966 fue designado jefe del departamento de astronomía y director del Observatorio Steward de la Universidad de Arizona. Desde su jubilación, ha vuelto, como él mismo dice, a su primera afición juvenil: la astronomía amateur. Bok completó recientemente el trabajo de preparación de la quinta edición de su clásica *The Milky Way* ("La Vía Láctea").

JOHN PETTITT, SOPHIE DUCKER y BRUCE KNOX ("Polinización submarina") son botánicos. Pettitt pertenece a la plantilla investigadora de la sección de historia natural del Museo Británico. Obtuvo su licenciatura y doctorado en el University College de Londres. Especialista en los mecanismos reproductores de las plantas, ha viajado a Australia en dos ocasiones para trabajar con los coautores del artículo. Ducker se jubiló de la Universidad de Melbourne en 1975, aunque continúa trabajando con estudiantes graduados. Berlinés de nacimiento, abandonó su país por oposición al régimen nazi, en la Nochebuena de 1938. Tres años más tarde se afincaba en Australia. Dedicó mucho tiempo y esfuerzo a la investigación de la acción antibiótica de los hongos del suelo. Su interés por las algas le movió a establecer el primer curso de botánica marina en Melbourne, en 1962. Knox enseña botánica en Melbourne. Natural de Escocia, estudió botánica en la Universidad de Edimburgo y se doctoró por la de Birmingham en 1962, ingresando en la Nacional Australiana en 1964.

IRVIN ROCK ("Percepción anortoscópica") es profesor de psicología en la Universidad Rutgers, de Newark, donde ha trabajado durante muchos años para el Instituto de Estudios Cognitivos. El próximo otoño se trasladará a la delegación de aquella universidad en New Brunswick, donde colaborará en el programa sobre la cognición. Graduado por el City College de Nueva York, se doctoró en 1952 por la New School for Social Research. Antes de ingresar en el claustro facultativo de Rutgers, en 1967, Rock había enseñado en la Universidad de Yeshiva.

MILTON J. FRIEDMAN y WILLIAM TRAGER ("Bioquímica de la resistencia a la malaria") iniciaron su colaboración hace unos años en la Universidad Rockefeller, en donde Friedman hizo su trabajo de licenciatura y Trager enseña y regenta el departamento de parasitología. Friedman es ahora biólogo investigador en el Instituto de Investigaciones sobre el Cáncer de la Universidad de California en San Francisco. Antes de graduarse por el Reed College trabajó sobre la bioquímica de la hemoglobina en el hospital clínico de la Universidad de Oregon y sobre genética bioquímica de organismos ecuarióticos simples, tanto en Reed como en la Universidad de Washington. Trager pertenece al claustro de la Universidad Rockefeller desde 1934, un año después de doctorarse por la de Harvard. Durante la segunda guerra mundial estudió la malaria en el Pacífico sudoccidental.

I. BERNARD COHEN ("El descubrimiento newtoniano de la gravitación") detenta la cátedra Victor S. Thomas de historia de la ciencia de la Universidad de Harvard. Obtuvo el título de doctor por dicho centro superior en 1947, siendo el primer norteamericano a quien se concedía un doctorado en este campo. Antiguo editor de *Isis,* ha presidido la History & Science Society y la Unión Internacional de Historia y Filosofía de la Ciencia.

Hace...

José M.ª López Piñero

... ciento cincuenta años

Francisco Javier Laso de la Vega y Orcajada fue nombrado catedrático del Colegio de Medicina y Cirugía de Cádiz. Nacido en Cartagena el año 1785, se había formado en el mismo Colegio gaditano, donde obtuvo sucesivamente los grados de licenciado en cirugía (1806), bachiller en medicina (1811), licenciado en medicina (1817) y doctor en cirugía (1825). Durante la Guerra de Independencia participó en varias acciones como cirujano militar y trabajó en los servicios sanitarios del puerto de Cádiz. Desde 1813, fecha en la que fue designado profesor sustituto de Medicina Práctica, desempeñó diversos puestos en el citado Colegio hasta llegar a catedrático. El otro gran escenario de la actividad científica de Laso fue la Sociedad Médico-Quirúrgica de Cádiz, de la que ocupó el cargo de secretario, así como el de director de su *Periódico,* que convirtió en la revista médica española más importante del primer tercio del siglo XIX.

Laso y la Sociedad Médico-Quirúrgica gaditana fueron los protagonistas fundamentales de la introducción en España de la nueva medicina anatomoclínica procedente de la escuela de París. La numerosa producción escrita de Laso incluye trabajos originales y traducciones, así como noticias y revisiones críticas acerca de las más importantes novedades médicas europeas. Se ocupó de los más diversos campos médicos, pero el núcleo de su obra fue lo que hoy llamamos medicina interna. El más notable de sus trabajos originales es, sin duda, una monografía sobre la fiebre amarilla (1821), detenido estudio de la clínica, la anatomía patológica y la terapéutica de esta afección. Incluye como apéndice unas "Inspecciones anatómicas", en las que Laso expone veintiséis casos propios, correspondientes a la epidemia de 1819. En cada uno de ellos ofrece la historia clínica y los datos de interés epidemiológico, un informe minucioso de la autopsia y un comentario, relacionando los resultados de ambas. Al final, ofrece un "Resultado general de las inspecciones anatómicas", del que citaremos, como ejemplo, la parte relativa a las lesiones gástricas: "Casi todos los estómagos se hallaban distendidos no sólo por el líquido que inundaba su cavidad, sino también por algunos, aunque pocos gases, generalmente fétidos: el exterior del ventrículo indicaba a veces un flogosis evidente por la inyección de los vasos que rampean en su superficie, y más frecuentemente se veía lleno de manchas lívidas, que en alguna ocasión, siendo profundas, se transparentaban al través de la prolongación del peritoneo que reviste su circunferencia: esta membrana serosa estaba toda muy inyectada en otro. En el momento advertimos muy frecuentemente que los vasos estaban voluminosos y llenos de una sangre líquida y oscura, con extravasión en alguno, hallándolo considerablemente lívido, y evidentemente gangrenado, en uno. En otros estaba bastante teñido del color ictérico. Abierto el estómago, se hallaba en su interior una cantidad de líquido, desde uno a tres cuartillos, de color negruzco, más o menos consistente, y como cortado, o semejante a los asientos del café; no pocas veces dicho líquido presentaba todo el aspecto de una sangre disuelta en cierta cantidad de agua. La superficie interior estaba constantemente ofendida. Desde el más simple flogosis, hasta el color negruzco y manchas evidentemente gangrenosas, había grados intermedios que confirmaban que esta víscera sufría la primera y más intensa acción de la causa morbífica. Generalmente un humor bruno y viscoso adherido a la superficie mucosa, que se podía raer y separar con el mango del escalpelo, barnizaba, por decirlo así, todo el interior. Desprendido éste se veían los vasos más o menos ingurgitados, manchas rojas de varia extensión, y otras lívidas, equimoseadas y negruzcas. Las membranas del ventrículo habían adquirido algún aumento en su espesor, fenómeno que fue bien considerable en cierto número. Hallamos el píloro sumamente estrecho en muchos, siendo de advertir que en éstos el vómito había sido constante y muy esforzado hasta la muerte. Es también de notar que en sólo dos individuos, en quienes se halló vacío el estómago, su orificio estaba tan contraído que con dificultad permitía el paso de una sonda".

En sus años juveniles, Laso había estado influido por un autor de transición como Philippe Pinel. Por el contrario, cuando realizó este trabajo sobre la fiebre amarilla, su mentalidad era ya la de un anatomoclínico *sensu stricto,* en la línea de la Charité parisina, encabezada por Jean N. Corvisart y René T. H. Laennec. No resulta extraño que fuera el introductor en España de la auscultación mediata como medio de diagnóstico habitual. André Mazet, médico francés amigo suyo, le envió un "pectorilocuo" a finales de 1820 o a comienzos de 1821, es decir, al año siguiente de la primera edición del tratado de Laennec. A comienzos de 1822, el *Periódico* de la Sociedad gaditana publicó un amplio resumen de este último, seguido de una lámina y del anuncio de un constructor local del nuevo instrumento: "El acreditado artista Don Joaquín de los Reyes ha construido dos pectoriloquos que compiten en perfección con el que le ha servido de modelo, remitido de París a Don F. Laso por el malogrado doctor Mazet. La Sociedad posee uno de aquellos, y el señor Reyes se ha ofrecido a construir este y cualesquiera otros instrumentos de su manufactura a precios cómodos". Seis años más tarde, el propio Laso pudo afirmar que se había convertido en algo habitual "en nuestras salas clínicas y en los aposentos de nuestros enfermos particulares".

Sin detenernos en sus otros trabajos sobre temas de medicina interna, anotaremos que tradujo el tratado de François Lallemand sobre la patología del encéfalo (1824-26), el mejor texto neurológico de la escuela de París.

Laso dedicó también gran atención a la terapéutica. Entre sus artículos consagrados a esta materia destaca quizás el titulado "Consideraciones sobre el uso médico de la quinina y cinconina" (1822), en el que resumió y comentó, por primera vez en España, los trabajos de Pierre Joseph Pelletier y Joseph B. Caventou, así como las investigaciones que François Magendie y otros autores habían realizado sobre dichos alcaloides. Asimiló estas aportaciones iniciales de la moderna farmacología desde los resultados de su propia experiencia, y se preocupó de difundir en su ambiente la obtención de los nuevos principios químicos activos. Resulta muy significativo que, dos años después, un farmacéutico miembro de la Sociedad publicara en el *Periódico* un método que había ideado "para elaborar con mayor economía la quinina". Laso introdujo asimismo el tratamiento iódico del bocio y se ocupó de otros recursos terapéuticos, entre ellos la electroterapia y las moxas de origen chino.

La higiene pública fue otra disciplina que mereció su atención. Trabajó sobre la vacunación antivariólica y fue un temprano difusor de la obra de un higienista tan significativo como Louis René Villermé. Bajo su directa influen-

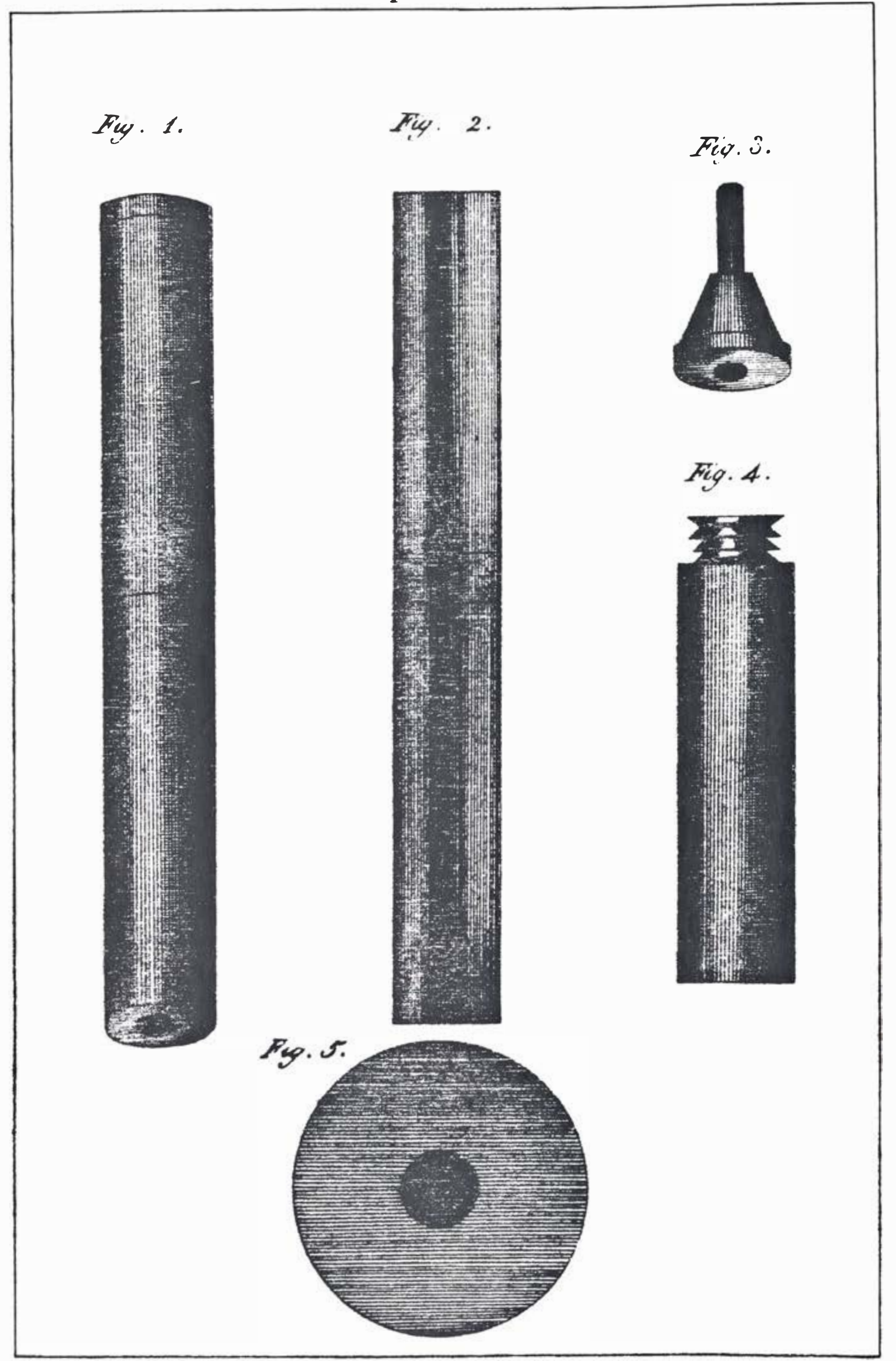


Lámina publicada a comienzos de 1822 en el **Periódico de la Sociedad Médico-Quirúrgica de Cádiz,** *en la que aparecen el primer estetoscopio o "pectorilocuo" construido en España. Francisco Javier Laso de la Vega introdujo en nuestro país la auscultación mediata, poco después de que Laennec publicara la primera edición de su célebre tratado sobre el tema*

cia, el *Periódico* de la Sociedad gaditana publicó una "Clave" (1820) para redactar topografías médicas e incluyó regularmente tablas con datos termométricos, barométricos e higrométricos de cada día del trimestre, así como una "constitución epidémica" del mismo, con exposición de las enfermedades dominantes y estadísticas de los fallecimientos, especificando causas de muerte, edades y sexos. La tradición ambientalista hipocrática y las incipientes tendencias medicosociales coincidieron, en suma, en este aspecto de su actividad.

Durante las dos décadas que siguieron a su muerte, en 1836, la orientación anatomoclínica que Laso había dado a la escuela de Cádiz fue continuada por autores como José de Gardoqui, Manuel José de Porto y Antonio Gracia Alvarez.

Energía solar fotovoltaica

La energía solar se utiliza por conversión térmica o por conversión directa en electricidad a través del efecto fotovoltaico. La tecnología de los dispositivos que realizan conversión directa, las células solares, ha registrado grandes avances

Luis Castañer

Para su subsistencia, la humanidad permaneció ligada de un modo casi exclusivo a la energía procedente del sol hasta el siglo pasado, cuando la innovación tecnológica precisó combustibles fósiles para su desarrollo. Esas fuentes energéticas parecen tener un plazo de agotamiento breve, por lo que el recurso solar vuelve a erigirse como vía alternativa deseable. No cabe pensar en una sociedad futura no electrificada, por lo que la dependencia, en una cantidad notable, de la energía solar tendrá que venir mediada por un proceso de conversión.

La cantidad de energía que se recibe del sol anualmente es del orden de $5100 \times 2,93 \times 10^{14}$ kilowatt hora, muy superior a la demanda mundial. La tecnología en curso se centra en su conversión en energía térmica y eléctrica. La conversión térmica se realiza a través de dos tipos de soluciones: a baja y a alta temperatura. Dentro de la primera solución se engloban el uso de la energía solar para el calentamiento del agua, calefacción de edificios residenciales y otros fines en la industria. La solución a alta temperatura consiste en un generador termospolar que sustituya a los actuales de carbón o nucleares para mover una turbina. Ambas técnicas se encuentran en un estadio de desarrollo bastante avanzado y es de esperar un rápido crecimiento en el número de watt que se produzcan durante los próximos años.

La conversión directa de la energía solar en electricidad se funda en el llamado efecto fotovoltaico que, en esencia, consiste en la producción de una fuerza electromotriz en un dispositivo, como resultado de la absorción de una radiación energética. El efecto fotovoltaico se produce en gases, líquidos y sólidos, siendo en los últimos donde se han registrado mayores rendimientos. Los dispositivos de estado sólido, diseñados para albergar dicho efecto, se llaman células solares. Deben transformar la radiación que procede del sol, equivalente a la emisión de un cuerpo negro a 5800 grados Kelvin de temperatura. Esta energía se distribuye sobre un rango de longitudes de onda que va desde 0,35 micrometros (un micrometro es la millonésima parte del metro, o la milésima del milímetro) hasta más de 3 micrometros. La célula solar, por tanto, debe ser sensible a ese intervalo de longitudes de onda y capaz de convertirla en energía eléctrica mediante un rendimiento de energía útil. La energía solar recibida en la superficie de la tierra al nivel del mar, y con el sol en el cenit, es de 1 kilowatt por metro cuadrado, aproximadamente.

La potencia eléctrica producida por la célula solar resulta de un proceso eficaz de interacción entre los fotones solares y el sólido que alberga el proceso. Consiguientemente, las propiedades fundamentales del material, así como sus posibilidades tecnológicas, condicionan y restringen el número de células solares posibles. El efecto fotovoltaico se produce cuando un fotón solar, o de cualquier otra fuente, es absorbido por el sólido y su energía utilizada en la creación de un par de portadores de corriente, que en un semiconductor son un electrón y un hueco. (Cuando un electrón absorbe la energía de un fotón y alcanza el nivel de energía conocido como banda de conducción, donde puede participar en la conducción eléctrica, deja tras de sí un "hueco", un lugar donde falta un electrón. Un electrón ligado en las proximidades puede moverse hacia el hueco, ocupándolo. Por tanto, electrones y huecos pueden moverse, y su movimiento conjunto constituye una corriente eléctrica.) Estos portadores tienen cargas eléctricas de signos opuestos, y para producir corriente en el exterior deben ser físicamente separados, separación que se produce mediante un campo eléctrico localizado en alguna zona del semiconductor.

El mecanismo de absorción depende del material que se utilice. En los materiales existe un valor mínimo de energía necesaria para que el fotón pueda crear un par de portadores, que se llama anchura de la banda prohibida, *Eg,* y es la diferencia de energía entre la banda de conducción y la banda de valencia. Los fotones que superan ese mínimo de energía pueden ser absorbidos en las proximidades de la superficie o bien pueden penetrar más en el interior del material. Este hecho se produce porque la absorción de los fotones sigue una ley exponencial gobernada por el coeficiente de absorción, que depende de la energía del fotón incidente de forma más o menos abrupta. Como quiera que la creación de un par de portadores se produce por la transición de un electrón de una banda de energía llena (banda de valencia) a una banda poco poblada (banda de conducción), la forma en que ocurra determinará la dependencia del coeficiente de absorción. Las transiciones de una banda a otra pueden ser directas o indirectas, según sea la posición relativa de las bandas de conducción y de valencia. Las transiciones directas producen un coeficiente de absorción abruptamente variable con la energía del fotón y las indirectas muestran una dependencia más suave.

Esta propiedad es intrínseca del material. Así, el silicio es un material en el que las transiciones son indirectas y el arseniuro de galio y el sulfuro de cadmio, materiales en los que las transiciones son directas. La energía necesaria para hacer que un electrón pase de un estado ligado a la banda de conducción se conoce también como *gap* de energía y es *Eg*; en este sentido, se habla del silicio como de un material de gap indirecto y, de gap directo, el sulfuro de cadmio y el arseniuro de galio. El gap de energía varía considerablemente de un semiconductor a otro, y se puede expresar en electronvolt. El silicio tiene un gap de energía de 1,12 electronvolt; por tanto, en un dispositivo fotovoltaico de silicio, un electrón necesita una

energía de 1,12 electronvolt para pasar a la banda de conducción; esta energía corresponde a una longitud de onda de 1,1 micrometros, que está en la región del infrarrojo del espectro.

En la ilustración de la derecha se representan esquemáticamente los coeficientes de absorción de varios materiales fotovoltaicos, así como la forma en que se absorben los fotones en el interior del semiconductor, para un coeficiente de absorción, α, grande o pequeño. Una conclusión interesante de dicho comportamiento es que, en los materiales de gap directo, el espesor de material necesario para absorber los fotones solares de energía superior al Eg correspondiente será mucho menor que el espesor preciso en un material de gap indirecto. Puesto que los espesores necesarios para los materiales de gap directo son del orden de algunos micrometros, si tecnológicamente el material en cuestión puede depositarse de forma homogénea y con buenas propiedades eléctricas en película fina, podrá construirse una célula solar de espesor muy reducido. Por el contrario, las células solares de materiales de gap indirecto, como el silicio, deben hacerse de espesores superiores.

De lo anterior se deduce que pueden haber numerosas soluciones, en lo referente a los materiales, para construir una célula solar. Teniendo en cuenta que el espectro del sol alcanza su máxima densidad de potencia en la región del ultravioleta, se precisará en general materiales de gap (Eg) no muy grande, por la sencilla razón de que se mostrarían transparentes a la mayoría de los fotones incidentes. Por otro lado, de un fotón de energía superior a Eg sólo se aprovecha Eg, lo que comporta que tampoco convengan materiales de Eg muy pequeños porque, si bien la mayoría de fotones producirían pares de portadores de corriente, la energía que se aprovecharía de cada fotón sería pequeña. J. J. Loferski hizo un cálculo considerando estos factores, entre otros, y demostró que el valor óptimo de Eg se situaba en torno a 1,5 electronvolt.

Otro punto característico de la célula solar es la llamada separación de carga. Un cristal de silicio dopado (impurificado) con fósforo es un semiconductor de tipo *n* porque tiene un exceso de electrones (exceso de carga negativa). Por el contrario, un cristal de silicio se puede dopar con boro, cuya valencia es tres; entonces el cristal será un semiconductor de tipo *p*, porque tiene un exceso de huecos cargados positivamente. El modo más habitual de crear un campo eléctrico dentro de un semiconductor es la formación de una unión *pn* mediante la introducción de átomos de impureza que creen un exceso de huecos, formando un semiconductor de tipo *p* y, análogamente, introduciendo átomos de impureza que creen un exceso de electrones, y de suerte que ambas zonas *n* y *p* estén en el mismo volumen semiconductor. En la unión o interfase entre el lado *p* y el lado *n* se crea un campo eléctrico de valor elevado, capaz, por tanto, de separar físicamente a portadores de carga de signo opuesto.

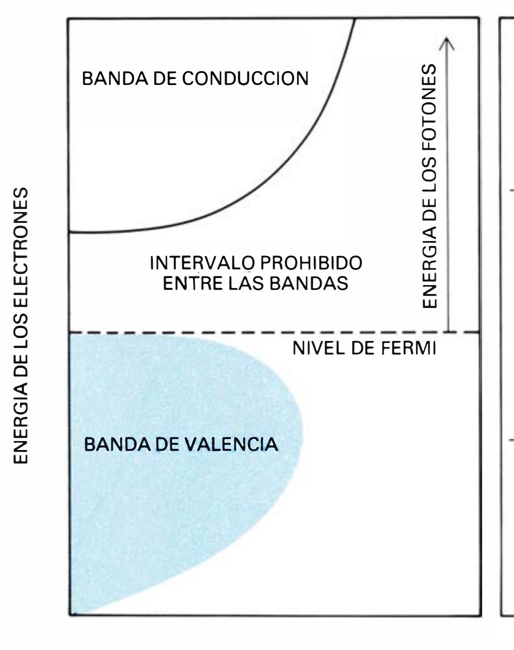
MATERIAL	Eg (eV)	ESPESOR (micras)
Silicio	1,11	150
Fosfuro de Indio	1,25	0,8
Arseniuro de Galio	1,40	2
Teluro de Cadmio	1,45	10
Sulfuro de Cadmio	2,43	1

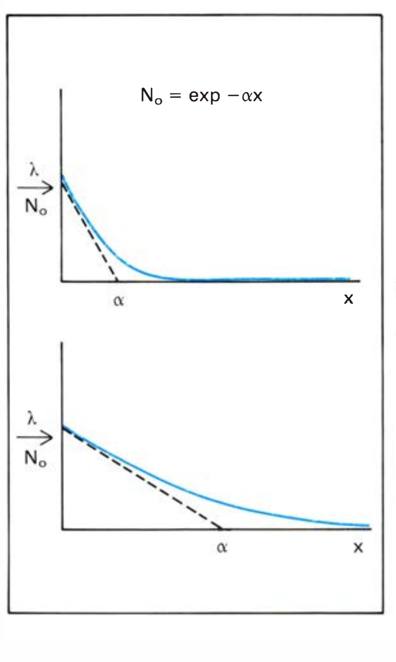
ESPESOR DE MATERIAL semiconductor preciso para absorber el 90 % de la energía del espectro solar para los fotones de energía mayor en el gap.

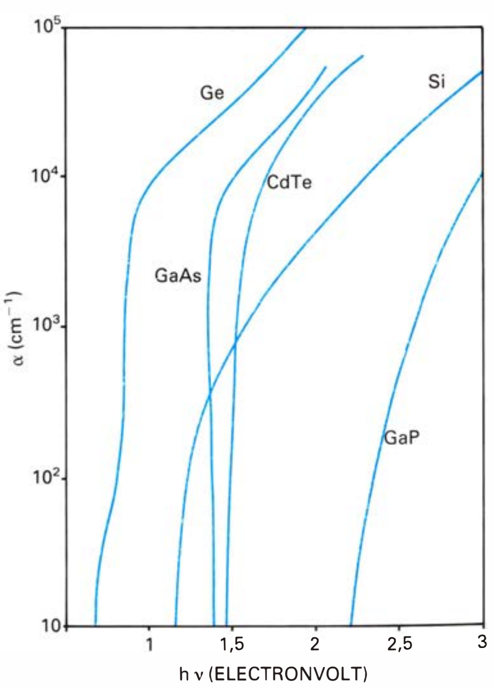
Una unión *pn* puede hacerse dopando *n* y *p* dos zonas del mismo material semiconductor. Recibe el nombre de homounión; por heterounión se entiende la unión *pn* fabricada mediante la unión de dos semiconductores diferentes, uno de los cuales sea de tipo *p* y otro de tipo *n*.

Se crea, asimismo, un campo eléctrico en uniones o barreras del tipo metal-semiconductor o, sus derivados, metal-óxido-semiconductor. En potencia, existen, pues, numerosos materiales fotovoltaicos con los que construir células solares en forma de homounión, heterounión o contacto semiconductor. En la actualidad el grado de desarrollo de la tecnología del silicio es el más elevado de todos los semiconductores.

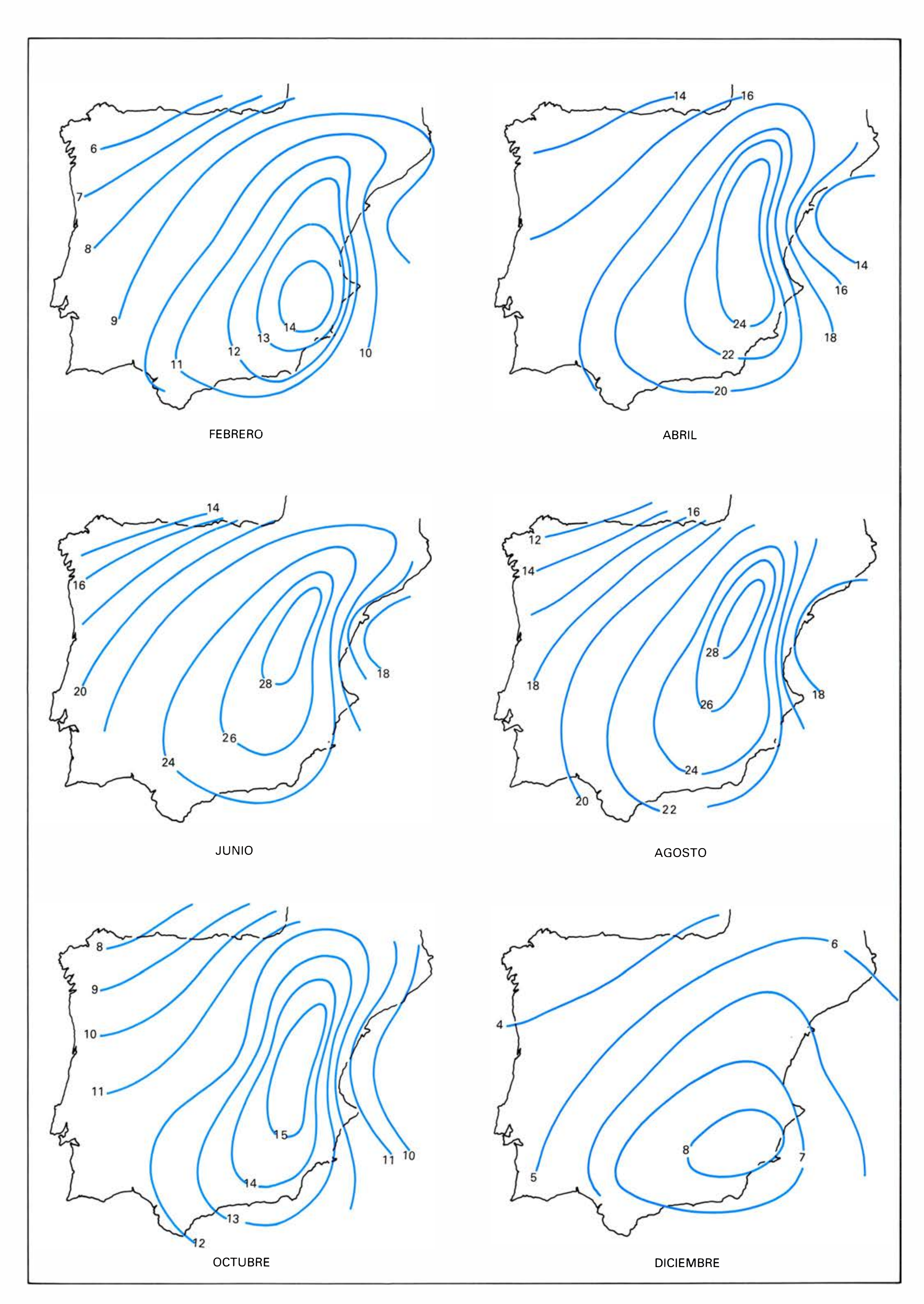
Dentro del propio silicio existen varias opciones. Consiste la primera en emplear material de pureza similar al que se utiliza para la fabricación de circuitos integrados y otros dispositivos





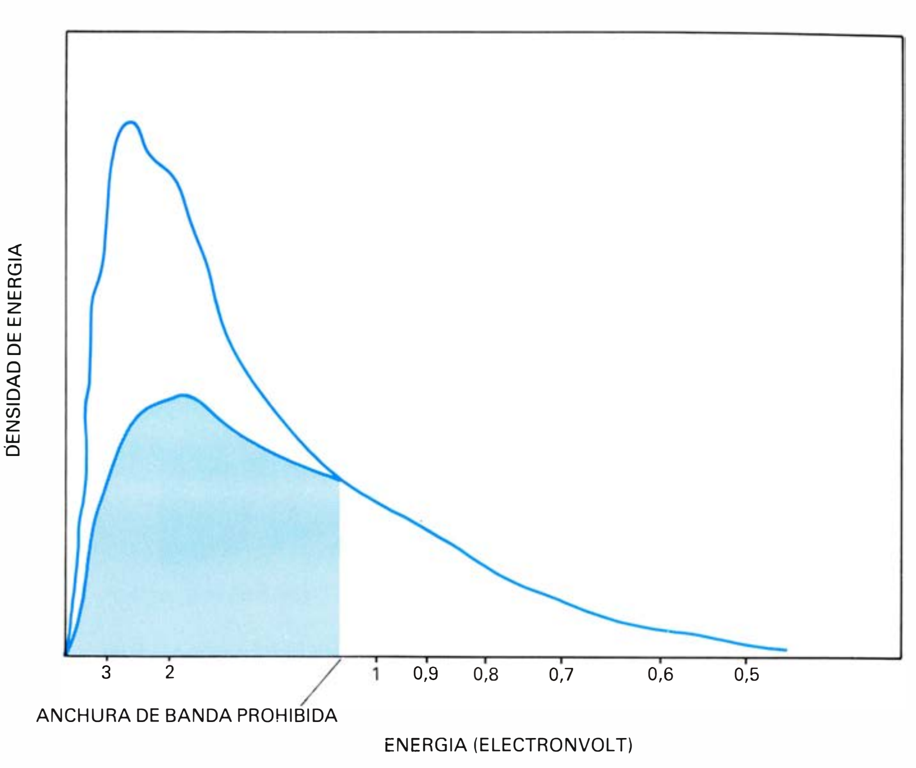


LOS FOTONES que llegan a la superficie de un semiconductor son absorbidos en función de la penetración. Para un valor grande del coeficiente de absorción α, ésta se produce en las proximidades de la superficie. Si dicho coeficiente es pequeño, los fotones incidentes son absorbidos en un espesor de material mayor. En la ilustración de la derecha se da la variación de los coeficientes de absorción de varios materiales en función de la energía del fotón incidente. Los materiales de gap indirecto, como es el caso del silicio, muestran una dependencia más suave que los materiales de gap directo.


6
7
8
9
11
12
13
14
10
FEBRERO
14
16
14
16
18
24
22
20
ABRIL
14
16
20
28
26
24
18
JUNIO
16
12
14
18
28
26
24
20
22
18
AGOSTO
8
9
10
11
15
14
13
12
11
10
OCTUBRE
6
4
5
8
7
DICIEMBRE

(silicio monocristalino) y mejorar el rendimiento de conversión, así como la automatización del proceso. Por tratarse de un material muy puro, su coste resulta elevado, lo que induce a exigírsele el máximo rendimiento posible. En esa línea se está trabajando ya.

Las llamadas células de silicio monocristalino sin concentración se fabrican con obleas de 75 y 100 milímetros de diámetro; constan, por lo regular, de una unión *pn,* donde la capa superficial es de tipo *n*. Entre la capa *p* y el contacto posterior se incluye otra capa *p* con un dopado más alto. La célula se completa con una rejilla colectora en la superficie expuesta a la luz y con una capa antirreflejo que reduce las pérdidas por reflexión en la superficie pulida de la célula. Esta configuración recibe el nombre de BSF (del inglés "back surface field", campo posterior); la corriente de cortocircuito que genera es próxima al máximo teórico calculado. Las mejoras en estudio se centran en la tensión de circuito abierto y en las pérdidas por resistencia. En células de este tipo se han conseguido rendimientos del orden del 18 por ciento, aunque las comercializadas en forma de panel sólo alcanzan un rendimiento del 12 por ciento.



ESPECTRO DEL SOL y parte empleada por un semiconductor de anchura de banda prohibida Eg, que es la diferencia de energía entre la banda de conducción y la banda de valencia. Para cada fotón dentro de la zona sombreada se utiliza Eg únicamente en la creación de portadores de carga, disipándose el resto.

Los sistemas de concentración con células de silicio utilizan concentradores estáticos o no estáticos. La concentración estática emplea espejos de Winston y la concentración no estática lentes de Fresnel y un sistema de seguimiento del sol. (El colector de Winston, llamado así en honor de su inventor, Roland Winston, es un reflector diseñado para reflejar una alta proporción de luz solar incidente sobre un área pequeña, incluso cuando el sol no está directamente sobre ella y, por tanto, no es necesario que el colector siga al sol.) Las células solares para estas aplicaciones deben tener pocas pérdidas resistivas por trabajar con densidades de corriente elevadas, así como estar ubicadas en una configuración adecuada para la disipación del calor.

Continuando dentro del silicio, la segunda posibilidad que existe es conseguir un silicio más barato mediante sistemas de producción diferentes del usual para rendir la oblea tradicional. (El sistema tradicional, o método de Czochraslki, consiste en el desarrollo de cristales perfectos de silicio a través de un proceso por el cual un germen de cristal gira al par que va saliendo lentamente de un crisol de silicio fundido. Si la temperatura y las velocidades de rotación y retirada se controlan con la necesaria precisión, se obtiene un monocristal perfecto.) Actualmente se trabaja en sistemas de crecimiento en cinta, según unos procedimientos que pueden abaratar grandemente el dispositivo una vez terminado.

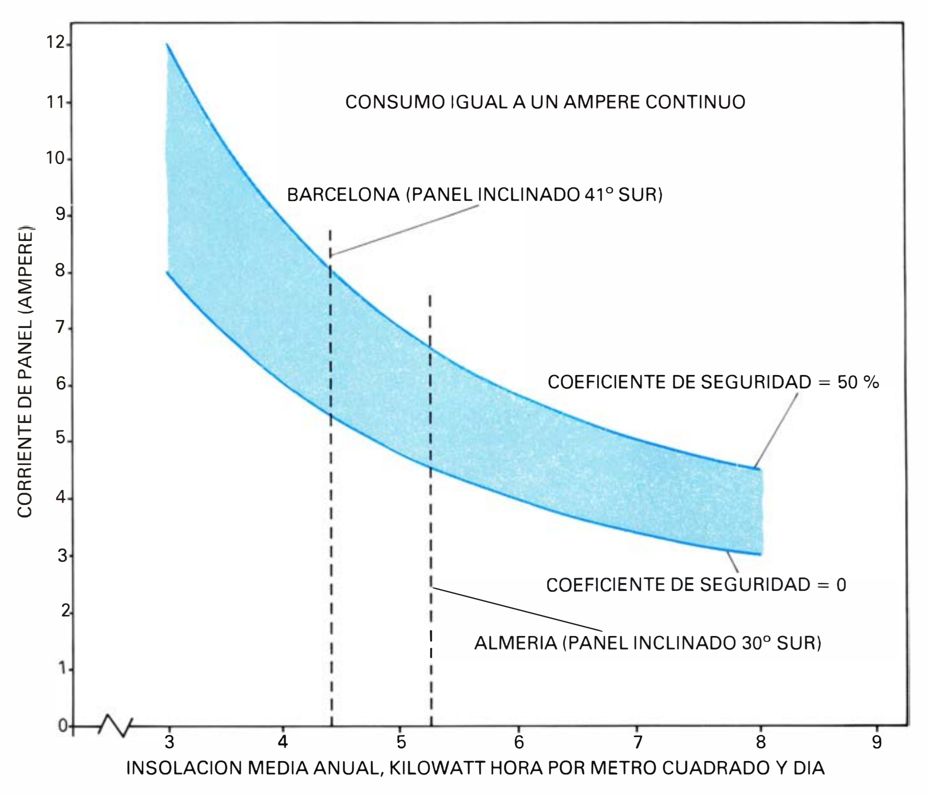
En la fabricación de células solares puede emplearse también silicio que no sea monocristalino, sino policristalino, formado por pequeños cristales elementales. Los rendimientos de conversión fotovoltaicos en este caso se sitúan hoy alrededor del nueve por ciento.

Puede producirse, por último, silicio semicristalino, cuyos granos miden típicamente varios milímetros de diámetro. Estos procesos, el crecimiento en cinta de alta pureza y la producción de silicio semicristalino y policristalino, tienen la ventaja de que pueden realizar células rectangulares con menos pérdidas de material que en el proceso convencional. Una ulterior reducción del precio del watt puede conseguirse mediante una mejora en los métodos de producción y automatización. Se han fijado los objetivos en conseguir un coste del watt pico de 2,86 dólares por watt en 1982, 0,7 dólares por watt en 1986 y 0,55 en 1990.

La segunda categoría de células solares es la de las células de película delgada o capa fina. Entran en esta categoría las soluciones tecnológicas que permiten espesores de unas micras, utilizando materiales de *gap* directo. En las previsiones de evolución del precio del watt pico solar, este tipo de dispositivos empezarán a ofrecer un coste comparable con las soluciones actuales a mediados de esta década, con una meta prefijada en los 0,1-0,3 dólares por watt hacia el año 1990.

En síntesis, una célula de película delgada está formada por una heterounión de dos materiales, uno de los cuales tiene un valor de Eg dentro del rango de rendimientos altos y el otro un Eg grande, actuando de ventana para la mayoría de los fotones procedentes del sol y sirviendo para crear el campo eléctrico que procurará la separación de carga. Las ventajas de que goza una estructura semejante son las siguientes: el espesor activo de la célula es muy reducido y puede depositarse sobre algún substrato barato que asegure un comportamiento mecánico adecuado. En la deposición de estos materiales se emplean procesos de estado sólido, como la evaporación, pulverización catódica, vaporización, etcétera. Como inconvenientes principales señalemos, por un lado, que las propiedades de los materiales capaces de formar heterouniones fotovoltaicas son mucho menos

RADIACION SOLAR recibida a lo largo de seis meses alternos en la península ibérica, en promedio. Las curvas de los mapas están rotuladas en megajoule por metro cuadrado. En cada área, la radiación solar varía con el clima, la hora del día y la estación del año. El autor detalla en el artículo las posibilidades que ofrece Almería, a 36 grados, 50 minutos y 18 segundos latitud norte. (Mapas cedidos por el profesor Manuel Puigcerver y Evangelina Oriol, de la Facultad de Físicas de la Universidad de Barcelona.)



CORRIENTE DE PICO DE PANEL, es decir, en condiciones estándar de 1 ampère para 100 miliwatt por centímetro cuadrado, necesaria para cubrir las necesidades de una instalación cuyo consumo equivalente resulte ser de 1 ampère. La zona tramada de la ilustración marca los límites entre un diseño sin considerar los factores de pérdidas y otro modelo más conservador que aplica un factor de seguridad de 1,5.

conocidas que las del silicio y, por otro, que el desarrollo de la tecnología de los mismos dista en general bastante de estar a un nivel en que la explotación industrial masiva sea viable. Asimismo hay que tener en cuenta que los materiales que se acoplen en una heterounión deben poseer ciertas características de compatibilidad desde el punto de vista cristalino, que, en el caso de no darse, introducen una cantidad elevada de enlaces rotos en detrimento del rendimiento.

Hasta la fecha, la célula de capa fina más desarrollada es la constituida por el sulfuro de cadmio y el sulfuro cuproso. El sulfuro de cadmio es un material de tipo *n* de Eg grande (2,43 electronvolt), que se deposita fácilmente en película delgada por evaporación, pulverización catódica y pirólisis por vaporización, cuyas propiedades eléctricas se controlan con bastante reproducibilidad. El sulfuro cuproso, que se consigue por intercambio iónico, tiene una anchura de banda prohibida (Eg) adecuada para un comportamiento fotovoltaico eficiente. Ha demostrado un rendimiento de conversión del nueve por ciento. Las variantes sobre este dispositivo se agrupan en dos tipos de modificaciones. La primera es la inclusión de cierto porcentaje de zinc en el sulfuro de cadmio, a fin de mejorar la compatibilidad cristalina. La otra alternativa consiste en el desarrollo de células entre un material de Eg próximo a 1,5 electronvolt y parámetros cristalinos muy próximos al sulfuro de cadmio y este último. En esta línea, los resultados más sobresalientes se refieren a las células de seleniuro de indio cobre/sulfuro de cadmio, así como otros materiales cuaternarios e incluso pentenarios. Los rendimientos actualmente conseguidos en capa fina están todavía por debajo del 5 por ciento. Hay que señalar que, en estos dispositivos, los rendimientos de los que se habla son resultados punta y, por tanto, superiores a los que se lograrían en una fabricación en serie. Los otros materiales que se usan como compañeros de heterounión del sulfuro de cadmio, que han producido células con rendimientos superiores al 5 por ciento, son el telururo de cadmio y el fosfuro de indio.

En otro orden de materiales, se asiste a una creciente actividad en la deposición de silicio amorfo, capaz de convertirse en conductor mediante la incorporación de hidrógeno. Una barrera de metal semiconductor con silicio amorfo hidrogenado ha demostrado un rendimiento fotovoltaico del 5,5 por ciento.

Entre las células fotovoltaicas citemos también las basadas en el arseniuro de galio, un material cuya tecnología se nos ha hecho familiar por su uso en la fabricación de diodos electroluminiscentes. En forma monocristalina, ha demostrado el rendimiento de conversión más elevado registrado hasta ahora, el 22 por ciento. Las células así construidas se comportan de un modo eficaz en sistemas con concentración, en razón del mantenimiento de sus propiedades a altas temperaturas. Con objeto de mejorar las características eléctricas de la homounión de arseniuro de galio en la capa de superficie, se introduce también cierta proporción de aluminio, formándose una heteroestructura de arseniuro de galio-aluminio/arseniuro de galio. El primero de los materiales tiene una Eg de 2,1 electronvolt, actuando como material ventana. Los rendimientos conseguidos con estos materiales en película delgada son sensiblemente inferiores a los conseguidos en forma monocristalina.

Finalmente, dentro de la panorámica que ofrece la conversión fotovoltaica existen sistemas más elaborados de conversión, como los proyectos de un mayor aprovechamiento del espectro solar por fragmentación y separación del mismo, de suerte que cada una de las partes incida sobre una célula de respuesta óptima en el margen correspondiente. Un sistema de este tipo que tenga un espejo selectivo y dos células, una de silicio y otra de arseniuro de galio-aluminio, puede alcanzar rendimientos superiores al 26 por ciento. El otro gran proyecto es el satélite fotovoltaico, cuya misión sería convertir la energía solar en un satélite en órbita geoestacionaria y enviarla a la tierra en la gama de las microondas, mediante un sistema de antenas emisor en el satélite y un sistema receptor en el suelo terrestre. Se trata, empero, de un proyecto cuya viabilidad y rentabilidad se hallan todavía en estudio.

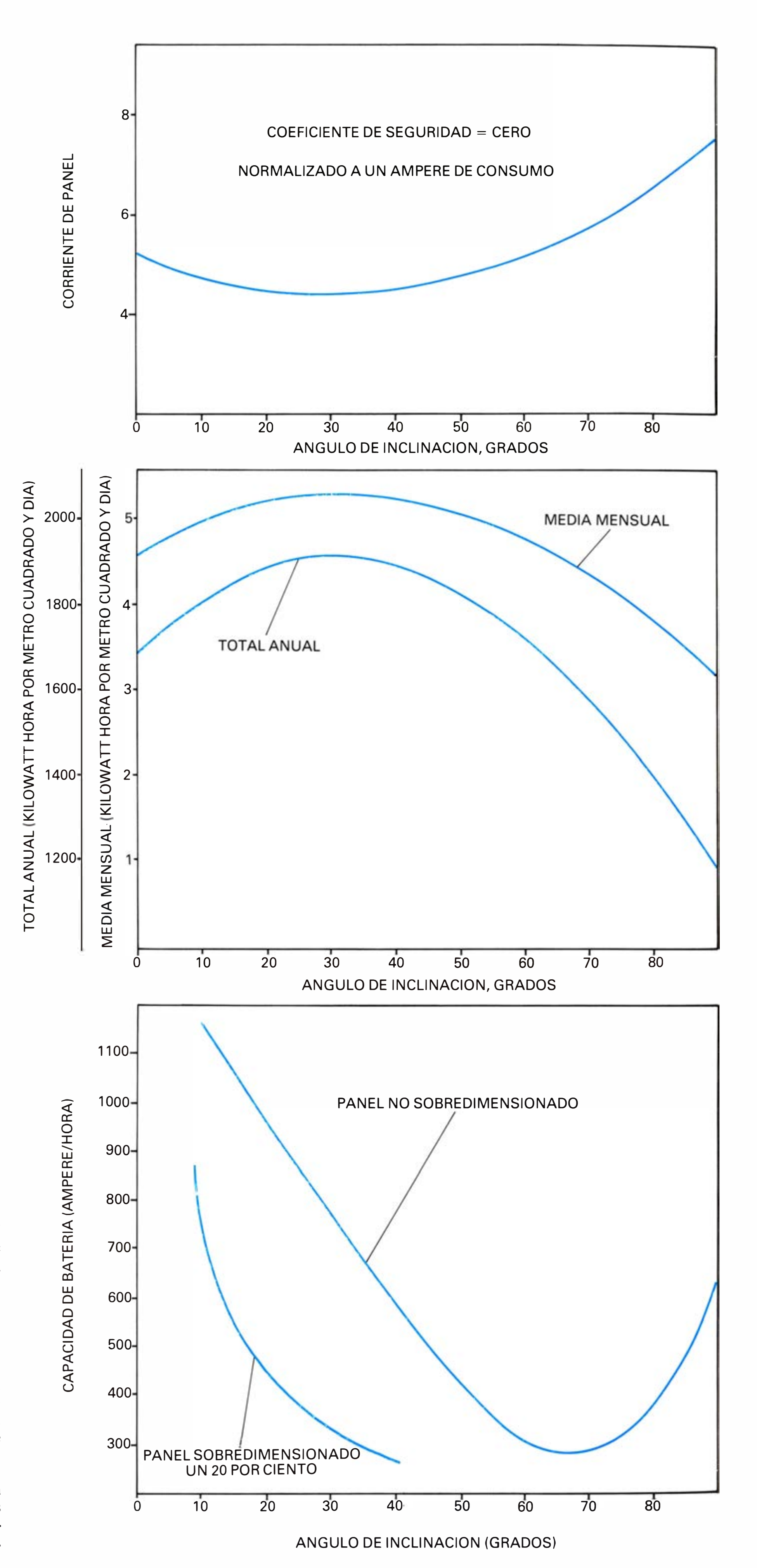
Un aspecto necesariamente complementario del dispositivo fotovoltaico es el relativo a lo que podría llamarse la ingeniería fotovoltaica, es decir, las cuestiones relativas al diseño y dimensionado de instalaciones para aplicaciones concretas, sin olvidar algunos aspectos de los diferentes bloques que componen el sistema, como el acondicionamiento de potencia y la adquisición de datos del comportamiento real de instalaciones.

Una célula solar aislada proporciona una potencia reducida. Por ejemplo, considerando un 12 por ciento de rendimiento y unas condiciones de iluminación de 100 miliwatt por centímetro cuadrado, una célula de 100 milímetros de diámetro suministraría 0,942 watt de potencia eléctrica continua, consistente

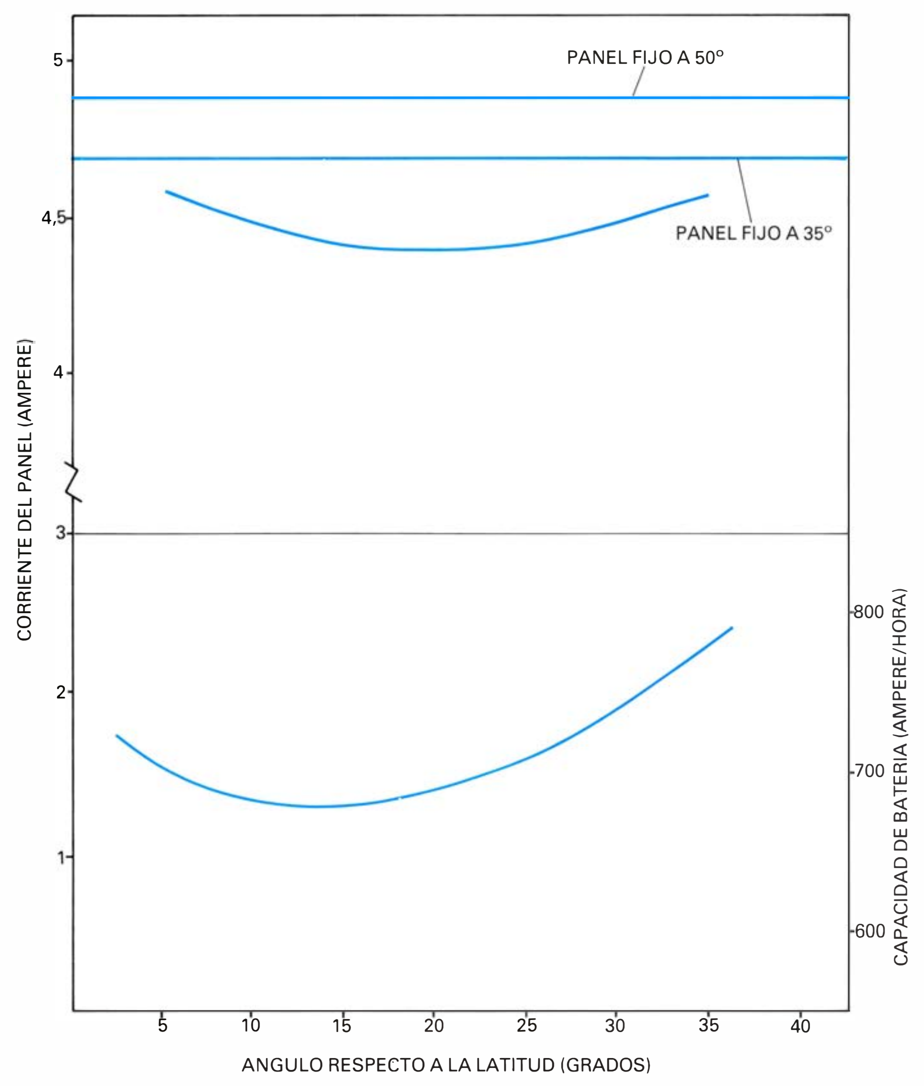
en una tensión de unos 0,5 volt y una corriente de unos dos ampère. A la vista de estos datos numéricos, los usos de las células aisladas se reducen a algunas aplicaciones de muy baja potencia. Para suministrar niveles de corriente y tensión mayores, las células solares deben asociarse en serie y en paralelo. En general, los paneles así constituidos constan de un conjunto de células en serie (de 18 a 36) para luego asociarse entre sí en paralelo.

Los paneles fotovoltaicos, una vez expuestos al sol, generan una potencia eléctrica continua, que puede aprovecharse directamente por los aparatos o equipos previstos para corriente continua, cuyo funcionamiento dependerá de la insolación que exista en cada momento. Por citar un ejemplo, el motor de corriente continua de una bomba de riego tendría velocidad variable a lo largo del día. Esta situación es, en algunos casos, perfectamente tolerable, mientras que en otras aplicaciones haría inviable la solución fotovoltaica. Hace falta, por consiguiente, un almacenamiento de energía que permita independizar la distribución temporal del consumo de la distribución temporal de la energía captada. Para que esta potencia sea aplicable, además, a sistemas de corriente alterna, se precisa de un sistema eficiente de conversión continua-alterna que satisfaga los requisitos de la carga. Los bloques fundamentales del sistema fotovoltaico serán, por consiguiente, el panel, el acumulador, el convertidor de corriente continua en alterna y las adaptaciones entre ellos.

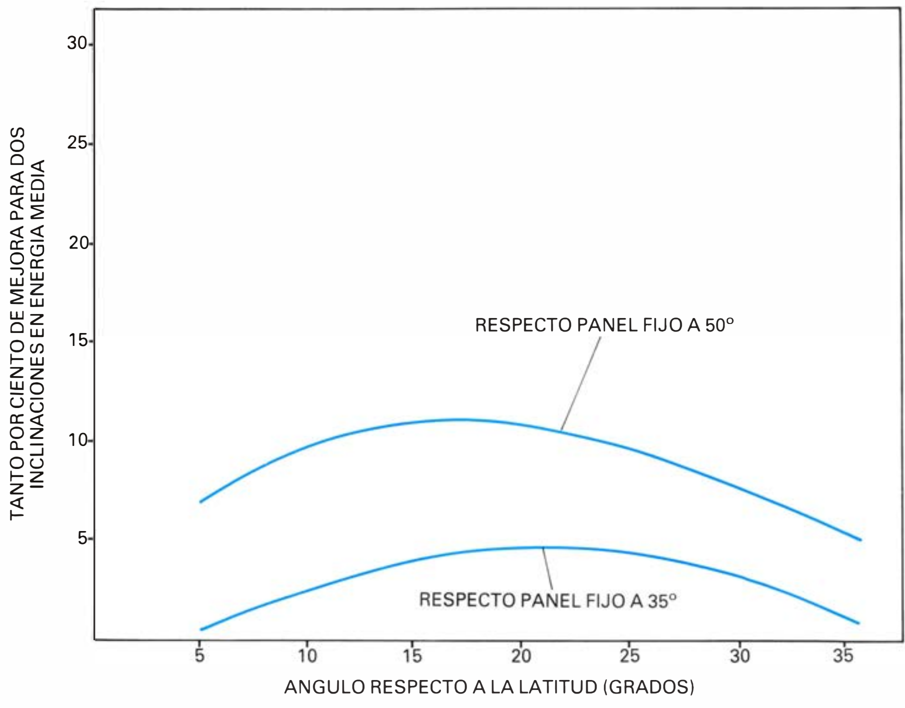
En el tamaño de una instalación fotovoltaica intervienen dos magnitudes fundamentales: la energía solar disponible en el lugar de la ubicación (la media diaria mensual y anual) y el consumo de la carga que se debe alimentar. Las instalaciones fotovoltaicas son, en su mayoría, fijas para usos residenciales o de comunicaciones, por lo que el diseño deberá hacerse sobre magnitudes medias, tanto de energía disponible como de consumo. El perfil de consumo de una instalación cualquiera puede reducirse al consumo continuo equivalente, es decir, a su valor medio anual.



EVOLUCION de la media mensual y del total anual de energía recibida en función del ángulo de inclinación del panel respecto a la horizontal, para el caso de Almería (*centro*). Arriba, la evolución de la corriente de pico de panel necesaria para garantizar un consumo equivalente a 1 ampère en función del ángulo de inclinación del panel. Abajo, la cantidad de ampère-hora precisos para cubrir los déficit anuales: cuando el panel no se sobredimensiona y cuando se dimensiona un 20 % del mínimo.



CANTIDAD DE CORRIENTE de pico de panel para un consumo equivalente de 1 ampère, si la inclinación del panel se modifica dos veces al año coincidiendo con los equinoccios, de modo que en invierno la inclinación sea un ángulo de 5 a 35 grados mayor que la latitud, y, en verano, de 5 a 35 grados menor.



AHORRO DE CORRIENTE DE PANEL que se consigue en el caso de Almería modificando la inclinación del panel dos veces al año (en porcentaje). El máximo se encuentra, en general, cuando en los meses de invierno el panel está inclinado un ángulo de 15 a 20 grados superior a la latitud e, inferior, en verano.

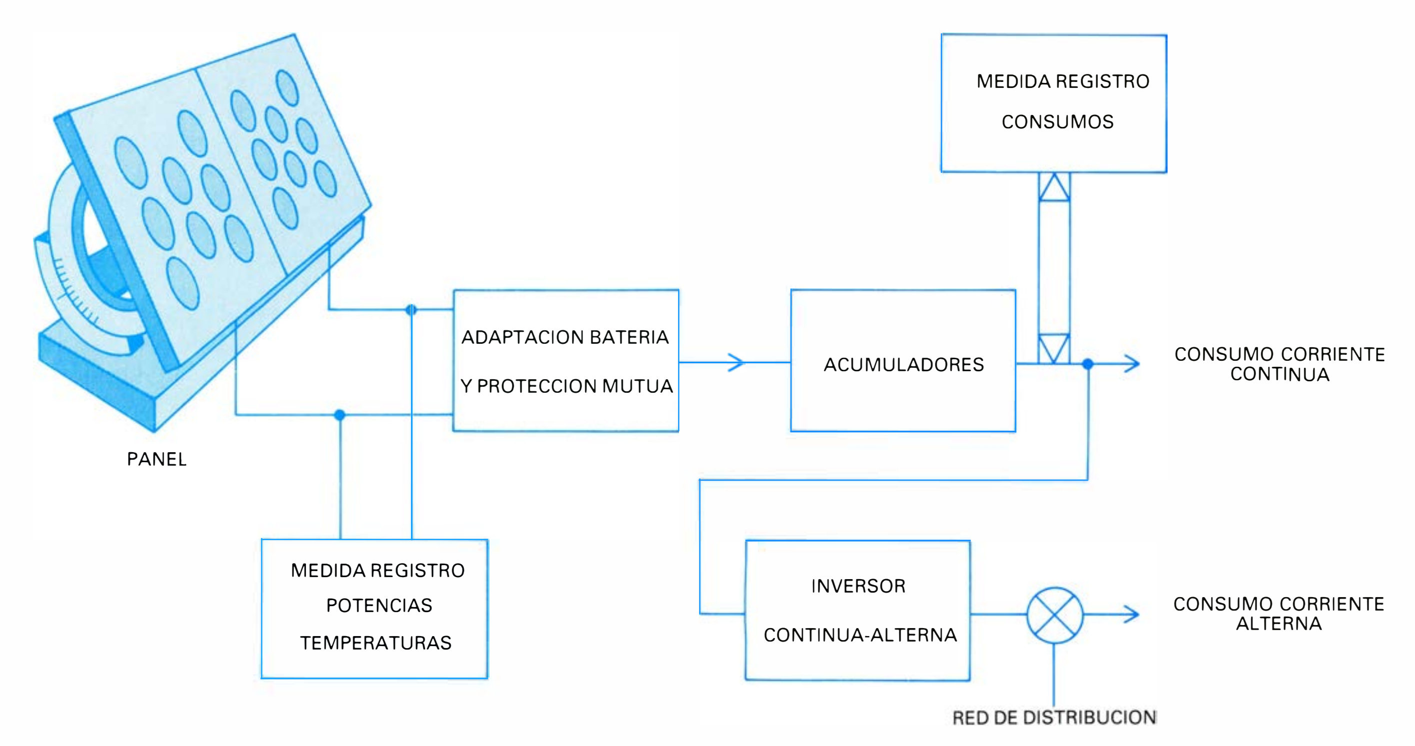
En cuanto a la energía media disponible, ésta es función de parámetros geográficos y astronómicos. La energía recibida en cada instante depende de la hora del día y del día del año. En la distribución anual de la energía disponible en la ubicación elegida intervienen también los factores del clima. Los parámetros más influyentes son la latitud, el ángulo horario y la declinación. Existen datos meteorológicos de insolación de numerosos puntos que son los promedios de bastantes años de observaciones, por lo que puede decirse que la energía recibida en un punto en un plano horizontal es una magnitud conocida de forma fiable.

De la observación de los datos para una cierta localidad se desprende que la energía disponible varía entre unos límites bastante amplios entre los meses de invierno y de verano. Si el plano de colección, en nuestro caso el panel fotovoltaico, se inclina cierto ángulo respecto a la horizontal en dirección sur pueden modificarse los límites de la variación, así como mejorarse la cantidad media anual de energía recibida.

Cálculos realizados considerando un plano colector fuera de la atmósfera terrestre demuestran que la energía recibida a lo largo del año es máxima, si el plano colector se inclina un ángulo cuyo valor coincida con la latitud del lugar de observación. Con datos reales de observaciones meteorológicas se comprueba que este resultado constituye una buena aproximación. El otro factor de interés es la variación, a lo largo del año, de la energía diaria media recibida en relación con el valor medio anual. Para ángulos de inclinación, con respecto a la latitud del lugar, grandes y pequeños, las fluctuaciones alrededor de la media son grandes.

Uno y otro factor, energía media anual y fluctuación alrededor de la media anual, influyen en la construcción de la instalación fotovoltaica. El primero, porque la cantidad de paneles fotovoltaicos necesarios es directamente proporcional al consumo continuo equivalente de la carga e inversamente proporcional a dicha energía media anual; y el segundo factor, porque la cantidad de acumuladores necesaria depende del déficit y del excedente de energía que se vaya produciendo con respecto a la media anual a lo largo de todo el año.

Sin considerar ningún factor de pérdidas, a las que se hace referencia más adelante, el número de paneles se suele determinar a partir de la corriente necesaria en condiciones de iluminación estándar. Conocido el valor medio



SISTEMA FOTOVOLTAICO compuesto por el panel, realizado mediante la combinación serie-paralelo adecuada a las necesidades de la instalación, un sistema adaptador de carga a los acumuladores, un conversor de continua en alterna y una serie de instrumentos de monitorización de toda la operación.

anual de la energía recibida, y por supuesto el valor de la corriente que entrega un panel en condiciones de iluminación estándar (de uno a dos ampère para iluminación de 100 miliwatt por centímetro cuadrado) se puede calcular el número total de ampère de panel necesario.

El diseño de la cantidad de acumuladores precisa hay que abordarlo partiendo de la base de que una instalación permanente debe tener una entrada de energía al año que compense exactamente el consumo anual y los posibles factores de pérdidas. Prescindiendo por el momento de los factores de pérdidas, es evidente que los acumuladores deberán poseer una capacidad suficiente para albergar los excedentes de energía que se produzcan y para cubrir los déficit en los períodos del año en que la energía captada esté por debajo de la media anual. Importa destacar que este papel de tampón de las baterías no se produce solamente siguiendo los ciclos nocturno y diurno, sino también siguiendo ciclos anuales. Esta circunstancia va a obligar a los acumuladores a trabajar períodos prolongados de tiempo en un estado bajo de carga.

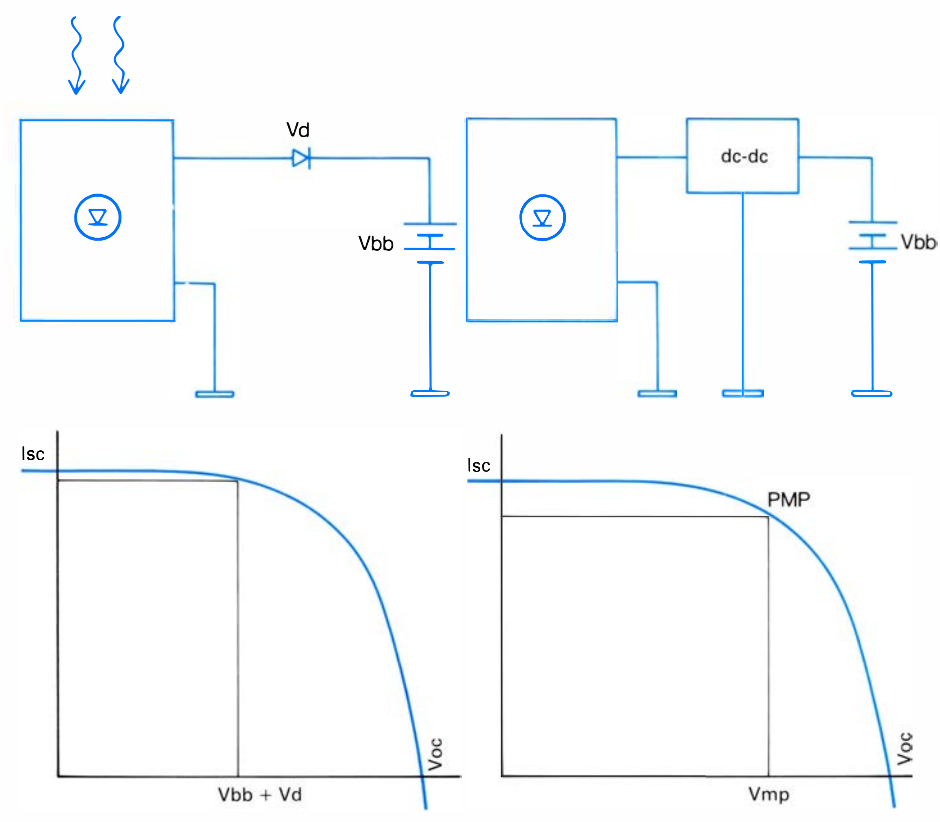
Calculando la cantidad de acumuladores necesaria para una instalación que tenga un consumo equivalente de 1 ampère, es decir, que gaste 24 ampère hora por día, con arreglo a los criterios anteriores, se puede concluir que dicha cantidad depende bastante de la inclinación que tenga los paneles respecto a la horizontal. En el caso de Almería, en el sureste de España, el ángulo de inclinación para el que se obtiene el mínimo de baterías necesarias no coincide con el ángulo de inclinación para el cual la energía captada a lo largo del año es máxima. Asimismo, cualquier factor de sobredimensión que se aplique para la corriente de panel repercute en una inmediata disminución de la acumulación necesaria.

La capacidad de la batería se calcula a veces considerando cierto número de días de autonomía de la instalación que, en general, aseguran un funcionamiento correcto, puesto que en la mayoría de los casos los paneles se sobredimensionan para tener garantías absolutas de operación fiable. La energía media anual y la fluctuación alrededor de la misma se modifican si cambiamos la inclinación del panel varias veces al año. Si se cambia coincidiendo con los equinoccios de primavera y otoño, es decir, en dos ocasiones al año, crece la media anual.

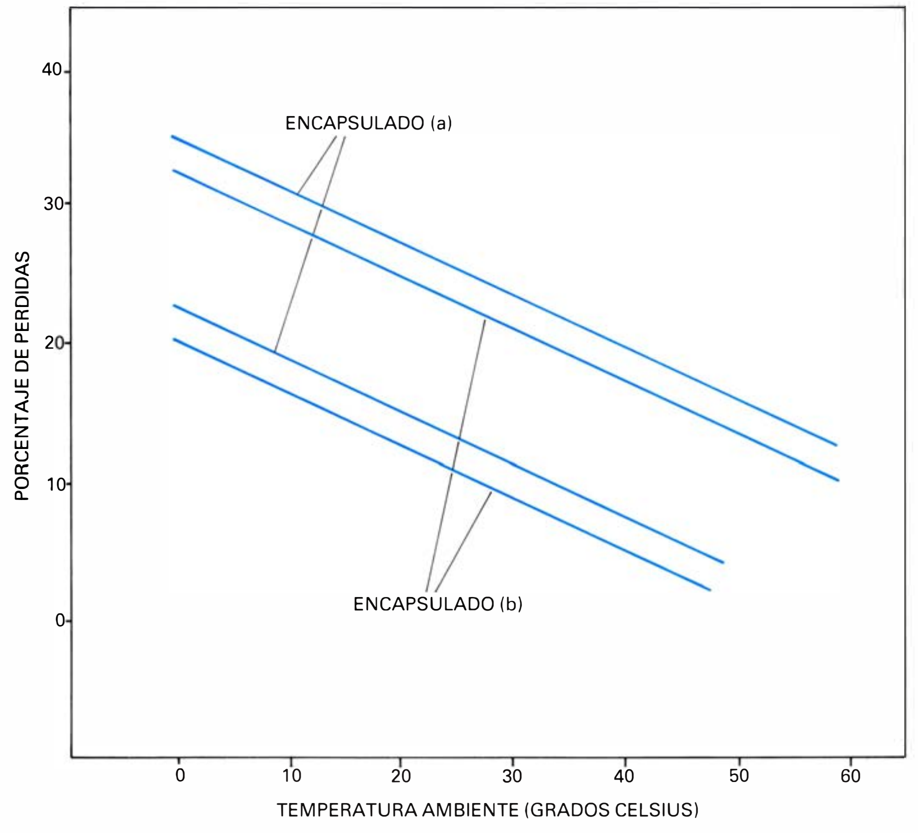
En la ilustración inferior de la página 12, para datos de Almería, se representa el valor medio anual de la energía captada en función del ángulo de inclinación con respecto a la latitud para los meses de verano y de invierno. En dicha figura, cinco grados representa que la inclinación del panel es cinco grados menor que la latitud del lugar en los meses de verano y cinco grados mayor en los meses de invierno. La energía recibida se ve aumentada en el mejor caso un 11 por ciento en relación con el panel fijo a 50 grados y un 3,8 por ciento respecto al panel fijo a 35 grados de inclinación. La capacidad de batería necesaria presenta también un mínimo para dos cambios al año con inclinaciones de 15 grados por encima y por debajo de la latitud del lugar. Hay que tener también presente que los paneles fotovoltaicos están comercializados en forma de módulos, por lo que el diseño definitivo de panel y baterías habrá de sufrir sucesivas aproximaciones.

Las consideraciones anteriores se han hecho sin tener en cuenta ningún factor de pérdidas, a excepción del propio rendimiento de conversión del panel. Pero hay múltiples causas que introducen pérdidas en el sistema. El panel fotovoltaico tiene una corriente y una tensión similares a las de la célula individual. Para una conexión en serie, la tensión es la suma de las tensiones individuales de las células y, la corriente, igual a la de una sola célula. Esta característica varía con la temperatura a razón de 25 miliampère por grado centígrado por centímetro cuadrado para la corriente de cortocircuito y de − 2,1 milivolt por centígrado por célula para la tensión de circuito abierto. Por su parte, la tensión de los acumuladores depende del estado de carga y de la temperatura a razón de − 5 milivolt por centígrado por elemento.

La conexión más usual panel-batería se da a través de un simple diodo, que impide la descarga de los acumuladores por el panel en los períodos de nubosidad o por la noche. Esta conexión im-



DOS FORMAS DE CONEXIONAR un panel fotovoltaico y una batería con inclinación mediante el sombreado de la potencia que se aprovecha en cada caso. Las abreviaturas designan lo siguiente: *dc-dc* = continua-continua; *Isc* = corriente de cortocircuito; *Vbb* = tensión de la batería; *Vd* = tensión en el diodo; *Voc* = tensión de circuito abierto; *PMP* = punto de máxima potencia y *Vmp* = tensión de este punto.



LA CONEXION DE UN PANEL A UNA BATERIA a través de un diodo obliga a perder una parte de la potencia que estaría disponible en la salida de la célula solar o del panel. La figura muestra una estimación de dicha pérdida en función de la temperatura ambiente y del estado de carga de las baterías (nominal B y bajo A). Asimismo, se han supuesto dos encapsulados diferentes que hagan que la diferencia de temperatura entre la célula y el ambiente sea de 15 grados en el encapsulado (*a*) y de 20 en el (*b*).

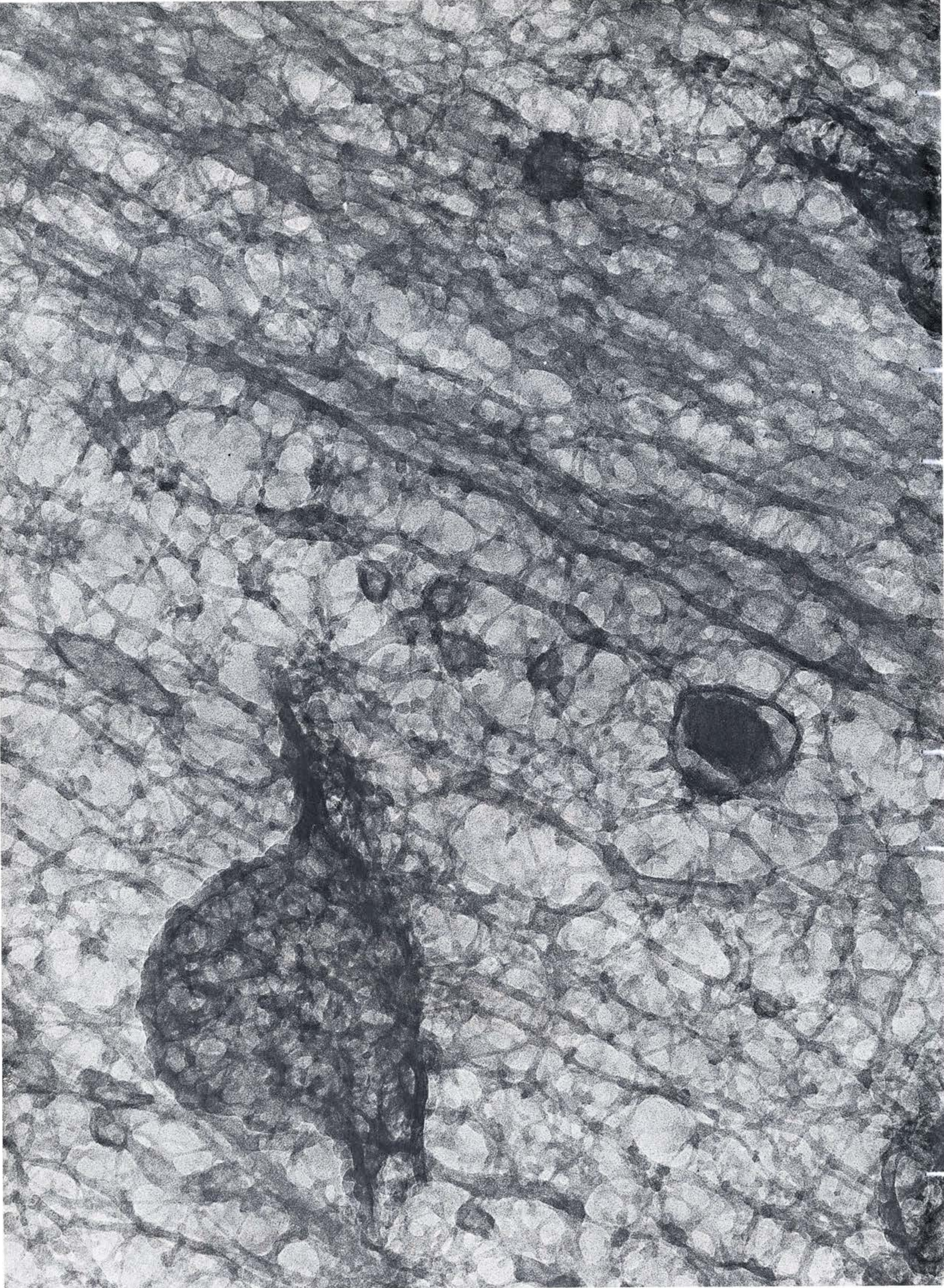
pone al panel cierta polarización, que no coincide con el punto donde la potencia es máxima (PMP). Por tanto, aunque el rendimiento nominal del panel sea del 12 por ciento, la potencia que se transmite a la batería puede llegar a ser considerablemente inferior. Esta adaptación es, además, función de la temperatura.

Se puede rebajar el valor de ese factor de pérdidas introduciendo, entre el panel fotovoltaico y la batería, un circuito electrónico llamado convertidor continua-continua, que obligue al panel a trabajar en el punto de potencia máxima, con independencia del nivel de iluminación y de la temperatura de trabajo. Los sistemas de adaptación panel-batería deben incluir también sistemas de protección para la carga excesiva de la batería. Esto se consigue mediante sistemas que detectan el máximo valor de tensión de la batería permisible e impiden, a partir de ese momento, la carga ulterior de la misma.

Aparece otro factor de pérdidas en los propios acumuladores, y es el factor de autodescarga (del 1,5 al 10 por ciento al mes), que debe ser cubierto por el sistema. Asimismo, la vida y la capacidad efectiva útil de los acumuladores depende del régimen de carga y descarga. En aplicaciones fotovoltaicas, las cargas y descargas son, en general, lentas, lo que favorece el comportamiento de los acumuladores. Sin embargo, las baterías deberán trabajar en estados de carga bajos durante cierto período de tiempo, con peligro posible de la vida de las mismas, sobre todo si se supera el límite máximo autorizado de descarga.

El último factor de pérdidas, aparte de la incertidumbre atmosférica y la opacidad progresiva del encapsulado de los paneles, es el rendimiento de conversión del inversor estático continua-alterna. El rendimiento de estos equipos es bastante alto para cargas fijas, es decir, si la carga está en sintonía con el circuito. En otras aplicaciones fotovoltaicas es preferible el uso de convertidores más versátiles que admitan cargas de diferentes tipos (fluorescentes y electrodomésticos).

Todos los factores de pérdidas citados repercuten en el diseño de las dimensiones de la corriente de panel necesaria y en el tamaño de la batería. Asimismo, habrá que contrastar los coeficientes de seguridad que se utilicen con resultados de instalaciones reales que permitan configurar modelos más ajustados. Las aplicaciones actuales van desde los relojes de pulsera hasta las centrales fotovoltaicas pasando por el campo de las comunicaciones.



El armazón celular

La gran resolución alcanzada con el microscopio electrónico de alta aceleración revela el retículo microtrabecular: un sistema de tenues filamentos que sostienen y desplazan a los órganulos de la célula

Keith R. Porter y Jonathan B. Tucker

Medio siglo atrás, se consideraba la célula cual una pequeña bolsa de líquido que encerrase unas cuantas estructuras en su interior. Con la aplicación del microscopio electrónico al estudio de la célula, se hizo evidente, en la década de 1940, que el citoplasma celular estaba organizado de forma más rigurosa. Presentaba múltiples fibras, vesículas rodeadas por una membrana y orgánulos. Se distinguían las mitocondrias (partículas que fabrican ATP, la moneda energético-química de uso e intercambio en la célula), los ribosomas (que coordinan la síntesis proteica de la célula) y el retículo endoplasmático y el aparato de Golgi (sistemas de vesículas que empaquetan las proteínas para su exportación fuera del territorio celular). Pero la microscopía electrónica no reveló gran cosa acerca de la naturaleza del medio, aparentemente no estructurado, en el que se suspendían los diferentes componentes celulares: la sustancia citoplasmática fundamental.

No obstante, ese armazón o sustancia fundamental presenta varias características que sugieren de una manera inmediata una estructura interna compleja. Dotada de propiedades viscoelásticas más propias de un gel que de un fluido, determina la forma de la célula, regenera componentes perdidos y dispone de la maquinaria molecular necesaria para el transporte directo de vesículas, gránulos y cromosomas. Además, los orgánulos no fluyen libremente, como lo harían en una solución carente de estructura, sino que mantienen una distribución radial alrededor del centro celular; lo que da a entender que están trabados por algún tipo de andamiaje interno.

En la década de los 70, el conocimiento de la estructura celular dio un paso importante con la comprobación de que la célula poseía una elaborada y dinámica red de elementos esqueléticos. Hay por lo menos tres sistemas citoesqueléticos químicamente distintos: los microtúbulos, los microfilamentos y los filamentos intermediarios. Estos elementos estructurales se asocian con la capacidad de movimiento y mantenimiento de la forma que se le reconoce a la célula. No se encontró, sin embargo, ninguna estructura en la sustancia citoplasmática fundamental que pudiera integrar y controlar el conjunto de esas diferentes redes de fibras, de suerte que pudiera explicar el comportamiento celular.

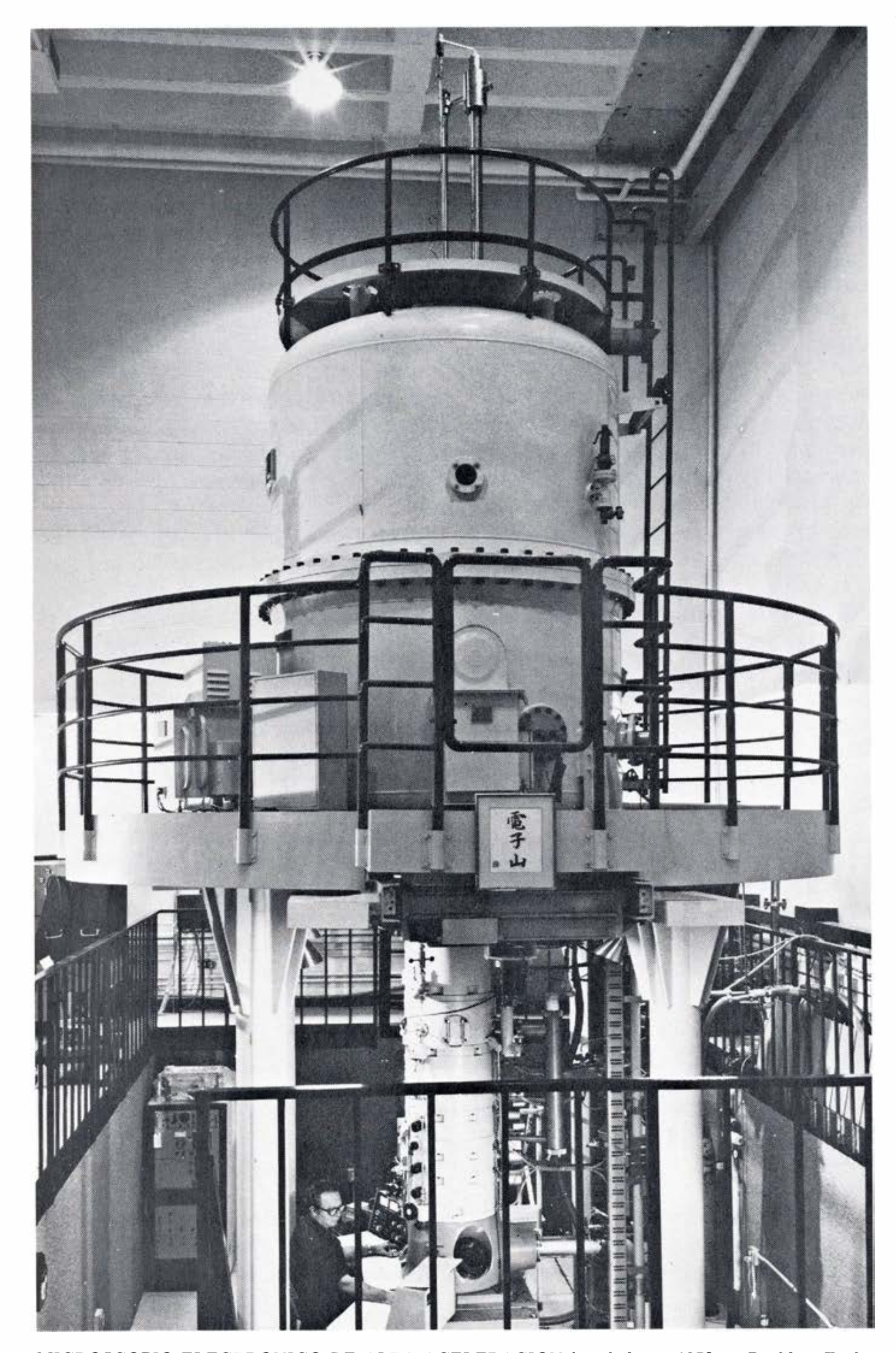
Una investigación más profunda obligaba al empleo de una herramienta analítica más poderosa que el microscopio electrónico convencional: el microscopio electrónico de alta aceleración. Aunque el primer instrumento de alta aceleración se construyó en 1947, hubo que esperar hasta los primeros años de la década de 1960 para que se desarrollaran modelos de investigación científica. Se debieron a Gaston-Léopold Dupouy, del Laboratorio de Optica Electrónica de Toulouse, y V. E. Cosslett, de la Universidad de Cambridge. Las versiones usuales del microscopio electrónico de alta aceleración miden unos nueve metros de altura, pesan más de 20 toneladas y aceleran electrones a través de una caída de potencial de un millón de volt, diez veces superior a la de los microscopios electrónicos tradicionales. Esta aceleración confiere a los electrones la energía cinética suficiente para atravesar muestras gruesas o incluso células enteras, de hasta varios micrometros de espesor. (Un micrometro equivale a una milésima de milímetro.) La imagen resultante, análoga a una radiografía, revela la estructura interna de la célula sin necesidad, como es el caso en aceleraciones menores, de seccionar la célula en cortes de un grosor inferior a los 0,2 micrometros.

La ventaja principal de la microscopía electrónica de alta aceleración es que proporciona información en profundidad, dando una visión sin paralelo de la organización tridimensional de las células. Además, la resolución y la profundidad de campo son excelentes por todo el grosor de la muestra. No obstante esas virtualidades, los biólogos han tardado en sacarle partido al microscopio electrónico de alta aceleración; y su principal uso en nuestros días se da, entre los metalúrgicos, para el estudio de metales en lámina fina. Se pueden contar con los dedos de la mano el número de esos aparatos empleados en la investigación biológica: varios en Europa y Japón y sólo tres en Estados Unidos.

En la primera mitad de la década de 1970, uno de los autores (Porter), a la sazón trabajando con Ian K. Buckley y John J. Wolosewick, empleó el microscopio electrónico de alta aceleración de la Universidad de Colorado en Boulder para investigar la estructura del armazón celular de la sustancia citoplasmática fundamental. Para el estudio se escogieron células desarrolladas en cultivo, pues tienden a extenderse sobre una superficie plana, y pueden así apreciarse en su integridad. Las células se

RETICULO MICROTRABECULAR, visible en esta electromicrografía de alta aceleración como una red de filamentos entrelazados. Corresponde a una zona minúscula, próxima al borde, de una célula cultivada de rata. El retículo sólo se aprecia a aumentos muy elevados, porque las microtrabéculas, es decir, cada uno de los filamentos, son finísimas: de unos seis nanómetros de grosor. (Un nanómetro es una millonésima de milímetro.) El retículo tridimensional sostiene, en esta imagen, un orgánulo celular, el lisosoma (*pequeño cuerpo oscuro a la derecha*), y entrecruza un haz de microfilamentos (*arriba*). El retículo también da soporte a un micropliegue: una formación que se extiende como una manopla desde la superficie celular (*abajo*). Los pequeños cuerpos redondeados situados en algunas intersecciones del retículo son ribosomas, orgánulos que intervienen en la síntesis proteica. La micrografía, realizada por K. L. Anderson, de la Universidad de Colorado en Boulder, aumenta las estructuras unas 100.000 veces.

cultivaban directamente sobre rejillas de oro de tres milímetros, cubiertas con una membrana plástica; se fijaban con glutaraldehído, se teñían y se sometían a una postfijación con tetróxido de osmio; se deshidrataban con alcohol y acetona. Se transferían luego a anhídrido carbónico líquido y se desecaban por el método del punto crítico; este método evitaba los efectos distorsionantes que causaba la tensión superficial, que provoca el colapso celular cuando el secado se hace al aire. Finalmente, la rejilla se introducía en la cámara portamuestras del microscopio, bajo vacío, y se observaba a gran aumento.

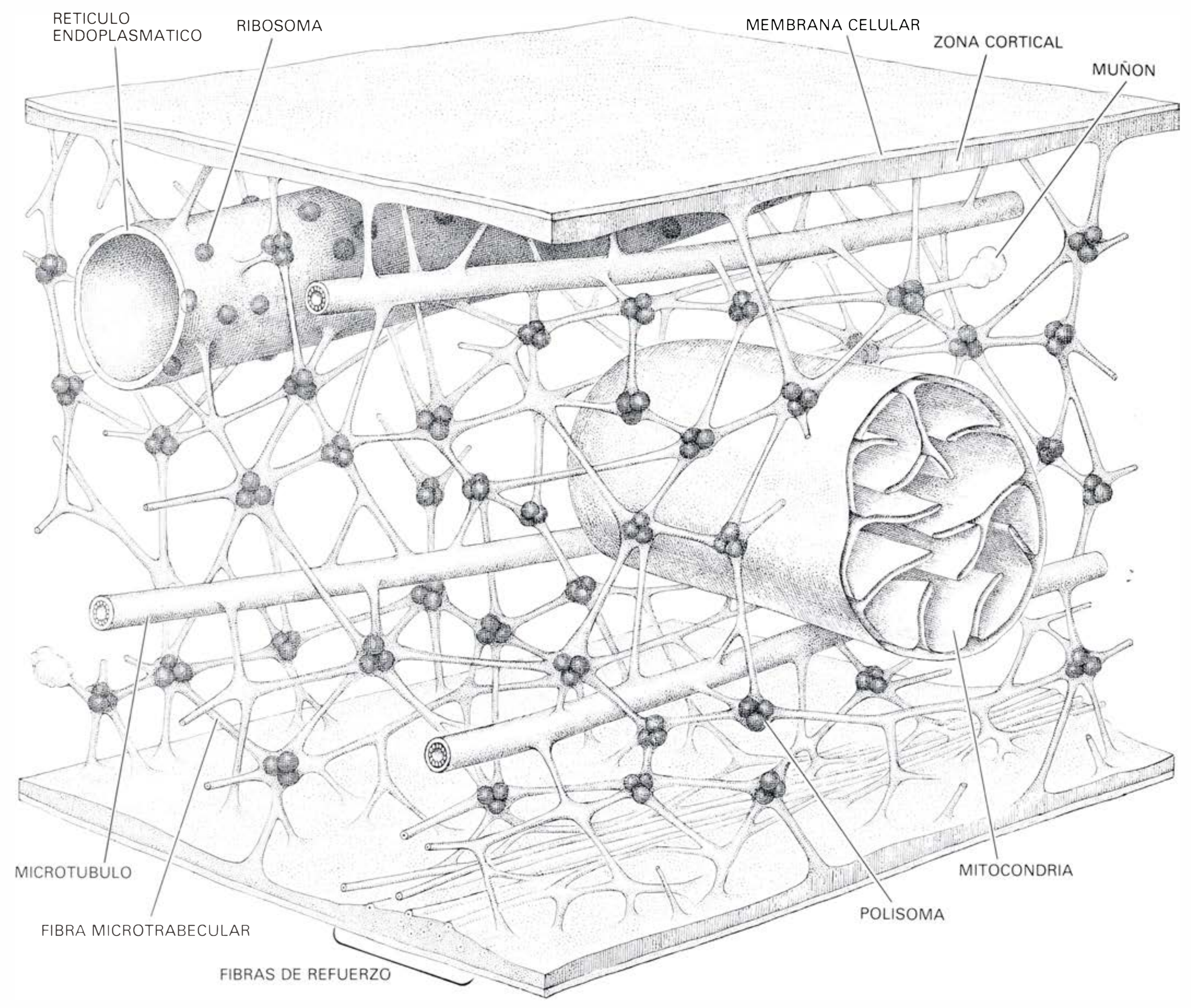


MICROSCOPIO ELECTRONICO DE ALTA ACELERACION instalado en 1973 en Boulder. En los Estados Unidos existen otros dos microscopios de este tipo destinados a la investigación biológica. El instrumento mide unos 10 metros de altura y pesa 22 toneladas. Para reducir las vibraciones se construyó sobre cimientos propios, independientes de los del edificio. Los electrones se inyectan en un tubo acelerador, encerrado en el gran tanque de presión situado en el centro, que aumenta su energía hasta un millón de volt. Estos electrones logran proporcionar imágenes de cortes gruesos, e incluso de células enteras, sin pérdidas importantes de resolución, lo que no se obtiene con electrones a potenciales más bajos. La columna óptico-electrónica debe ser lo suficientemente grande como para acomodar las potentes lentes electromagnéticas requeridas para desviar y enfocar los electrones de alta energía. Dada la intensidad de los rayos X generados donde los electrones chocan con una superficie, la columna del microscopio y su base están apantalladas para proteger al operador. La inscripción japonesa reza "montaña electrónica".

Al examinar zonas, no muy extensas, del armazón citoplasmático próximas al borde celular, los investigadores de Boulder advirtieron un retículo irregular y tridimensional de esbeltas fibras por todo el citoplasma, que llegaba hasta la membrana celular externa. Los filamentos entrelazados parecían sostener los distintos sistemas y orgánulos celulares. Además, los elementos citoesqueléticos mayores, como los microtúbulos y los microfilamentos, aparecían recubiertos por un material matriz imbricado en los filamentos del retículo, evidenciando que los elementos mayores formaban un todo con el retículo y, esencialmente, estaban suspendidos en él. El retículo, que recordaba la estructura trabecular del hueso esponjoso, se le bautizó con el nombre de retículo microtrabecular. A partir de entonces se ha observado dicha estructura reticuliforme en la sustancia fundamental de todas las células eucariotas.

En 1976, *The American Journal of Anatomy* publicó la primera descripción completa del retículo microtrabecular; la reacción de no pocos investigadores fue de escepticismo. Alegaban que las duras condiciones requeridas para fijar y teñir las células para microscopía habían provocado la condensación, alrededor de diversos orgánulos y fibras, de las proteínas citoplasmáticas en solución, provocando una falsa imagen de matriz estructurada. El grupo de Boulder se concentró con ahínco en la demostración de su conclusión, aplicando diversas técnicas de fijación y deshidratación en la preparación de células enteras para microscopía electrónica de alta aceleración. La sustancia fundamental presentaba siempre la morfología reticular. Además, soluciones ricas en proteínas sometidas a los mismos tratamientos no revelaban ningún artefacto de fijación que se pareciera al retículo microtrabecular.

No se niega que el retículo varíe algo en determinados pormenores de su morfología según el método de fijación a que se recurra; por ejemplo, en células fijadas en glutaraldehído, las microtrabéculas pueden tener un grosor superior a los 10 nanómetros en los extremos, y sólo de dos o tres en su parte media. (Un nanómetro es la millonésima parte de un milímetro.) Si las células se fijan en milisegundos por congelación en propano enfriado a la temperatura del nitrógeno líquido (– 196 grados Celsius), las microtrabéculas son más gruesas y presentan un diámetro más constante, de unos 15 nanómetros. Esta diferencia en la morfología



MODELO DE RETICULO MICROTRABECULAR, aumentado unas 300.000 veces, deducido a partir de centenares de imágenes de células cultivadas observadas con el microscopio electrónico de alta aceleración. El modelo representa la relación que guardan los filamentos microtrabeculares con los demás componentes del citoplasma celular: la sustancia de la célula que queda por fuera del núcleo celular. En el modelo, las microtrabéculas sostienen las estructuras elongadas del retículo endoplasmático (el sistema de canales interconectados donde quedan secuestradas algunas de las proteínas fabricadas por la célula), las mitocondrias (los orgánulos que fabrican ATP, el combustible universal de la célula), los microtúbulos (complejas fibras con múltiples funciones estructurales dentro de la célula) y los microfilamentos enterrados en la zona cortical celular (la capa de material situada justo por debajo de la membrana celular externa). En algunas de las intersecciones del retículo microtrabecular hallamos los polisomas: grupos organizados de ribosomas.

del retículo se debe, verosímilmente, a que el glutaraldehído es un fijador bastante lento, que requiere segundos para ejercer su pleno efecto, mientras que la congelación es prácticamente instantánea. Así, durante la fijación con glutaraldehído, las microtrabéculas tienen tiempo de contraerse o alargarse, alterando su diámetro y su relación entre ellas y con los orgánulos citoplasmáticos.

Pero si el retículo microtrabecular existe, ¿por qué no se ha apreciado antes en las micrografías electrónicas tradicionales? La verdad es que lo había sido, aunque de una manera tan indefinida que impedía su reconocimiento como tal. En cortes finos, las microtrabéculas aparecen como estructuras tenues y amorfas, que fácilmente se toman por artefactos; la organización de los filamentos en retículo no aparece por ningún lado. Cabe atribuir parte de la confusión a la finura del corte, que no permite una observación nítida de las interconexiones de los elementos del retículo. Además, la resina epoxi, que sirve de material de inclusión para cortes finos, tiene el inconveniente de enmascarar las estructuras celulares que se comporten, en lo relativo a la dispersión electrónica, de una manera similar a la misma resina.

Para resolver este problema, el grupo de Boulder empezó por preparar cortes finos de células incluyéndolas en un medio hidrosoluble: polietilenglicol. Una vez cortado el material para microscopía electrónica, se disuelve y se desecan los cortes por el método del punto crítico para evitar su contracción. Al observar estas secciones con un microscopio electrónico convencional se aprecia una morfología reticular muy parecida a la que se distingue, en células enteras no incluidas, con el microscopio electrónico de alta aceleración. Otras técnicas de preparación menos agresivas, como el método de criofractura empleado por John E. Heuser, de la Universidad de Washington, proporcionan una estructura reticular similar. Así pues, aunque algunos aspectos del retículo microtrabecular han resultado ser artefactos (y hay que puntualizar que los rigurosos métodos de prepa-

ración de la microscopía electrónica hacen que, hasta cierto punto, todo lo observado bajo el microscopio electrónico sea un artefacto), hay pruebas suficientes que indican la existencia real de un armazón estructurado en las células.

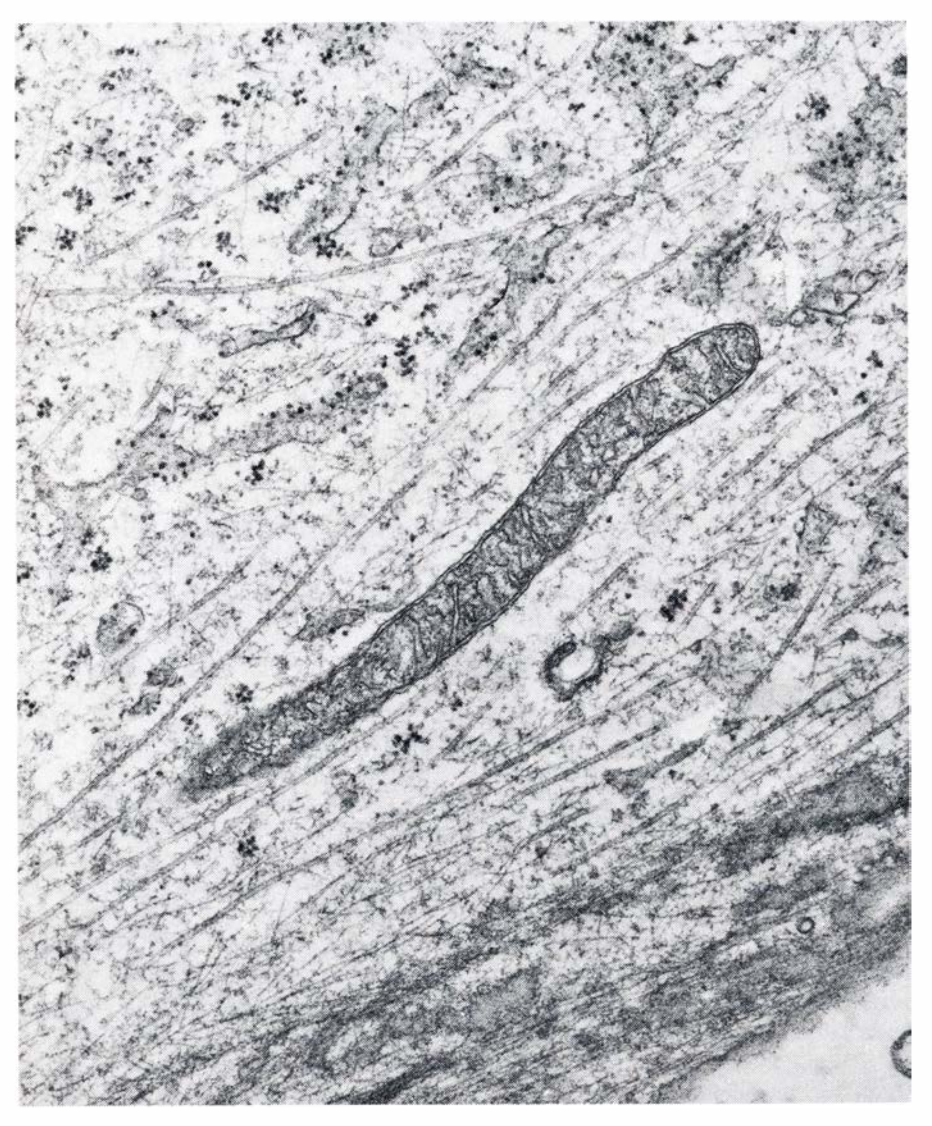
Tal vez la prueba más convincente de la existencia del retículo microtrabecular sea la variación de su estructura en respuesta a cambios de la forma o del ambiente celulares. Por ejemplo, a una temperatura baja (cuatro grados Celsius), las células cultivadas devienen esféricas, su configuración de menor energía. Este cambio es paralelo a la desorganización secuencial de las estructuras esqueléticas internas de la célula. Primero se disgregan los microtúbulos y, luego, los microfilamentos. Por último, se deforma el retículo microtrabecular, aunque sin descomponerse totalmente. Algunas microtrabéculas cortan sus conexiones con el retículo y se retraen formando diminutos "muñones"; se crean así amplios espacios en el retículo, que permiten el movimiento browniano (difusión libre) de los orgánulos citoplasmáticos, un fenómeno observado en células enfriadas, pero no en las que están a temperaturas fisiológicas. Si la célula vuelve a adquirir la temperatura corporal (37 grados Celsius), durante tan sólo cinco minutos, asistimos a la reestructuración del retículo en su forma primitiva.

Efectos similares sobre la estructura del retículo microtrabecular pueden provocarse con la citocalasina B, un fármaco que inhibe una amplia variedad de movimientos celulares. Después de diez minutos de exposición al fármaco, el retículo se torna muy basto: las microtrabéculas aumentan de grosor y se alargan, lo que implica, en correspondencia, que los espacios intertrabeculares se ensanchen. Al poco de suprimir el fármaco, vuelve el retículo a su estructura normal, finamente dividida. Otros agentes y condiciones (presión osmótica excesivamente alta o baja, niveles anormales de iones específicos, como el magnesio o el calcio, y presencia de inhibidores metabólicos) inducen también cambios estructurales únicos y reversibles en el retículo. Sería difícil explicar estas amplias fluctuaciones morfológicas si la estructura en cuestión se tratara de un mero artefacto de la fijación celular.

¿Cómo incide el retículo microtrabecular en la estructura y función celulares? Parece claro que el retículo divide la célula en dos fases: la polimerizada, fase rica en proteínas del propio retículo, y la líquida, fase rica en agua que llena los intersticios del retículo (los espacios intertrabeculares). Con un contenido en agua del 50 por ciento, esta construcción parece dar al citoplasma su consistencia gelatinosa.

La realidad de ambas fases fue demostrada experimentalmente por el grupo de Boulder (al que se había sumado Karen L. Anderson). En los ensayos se manipulaba el contenido de agua de las células. Sumergíamos las células en soluciones que contenían concentraciones altas o bajas de moléculas de azúcar, lo que provocaba la salida o entrada de agua por ósmosis. Al examinar las células al microscopio electrónico de alta aceleración, se vio que los espacios intertrabeculares se expandían si el agua entraba en la célula y se encogían cuando salía el agua. Si se extraía de las células el agua y se reemplazaba poco después, sobrevivían. Esta observación puede explicarse por la capacidad del retículo para preservar la viabilidad de las células, protegiéndolas de las fluctuaciones normales de su contenido hídrico.

El retículo microtrabecular quizá desempeñe también alguna función en la organización de los enzimas del citoplasma. En 1929, Rudolph Peters, de la Universidad de Cambridge, sostuvo, basándose en pura lógica, que las intrincadas reacciones bioquímicas que se sabía ocurrían en el citoplasma estaban organizadas de forma sumamente rigurosa como para depender de las colisiones fortuitas de los enzimas con los sustratos (las sustancias que aquéllos transforman), que difunden a través de un medio carente de estructuras. Postulaba la existencia de "una tenue red por cuya acción se coordinan las actividades enzimáticas celulares". El razonamiento de Peters era clarividente. Dis-



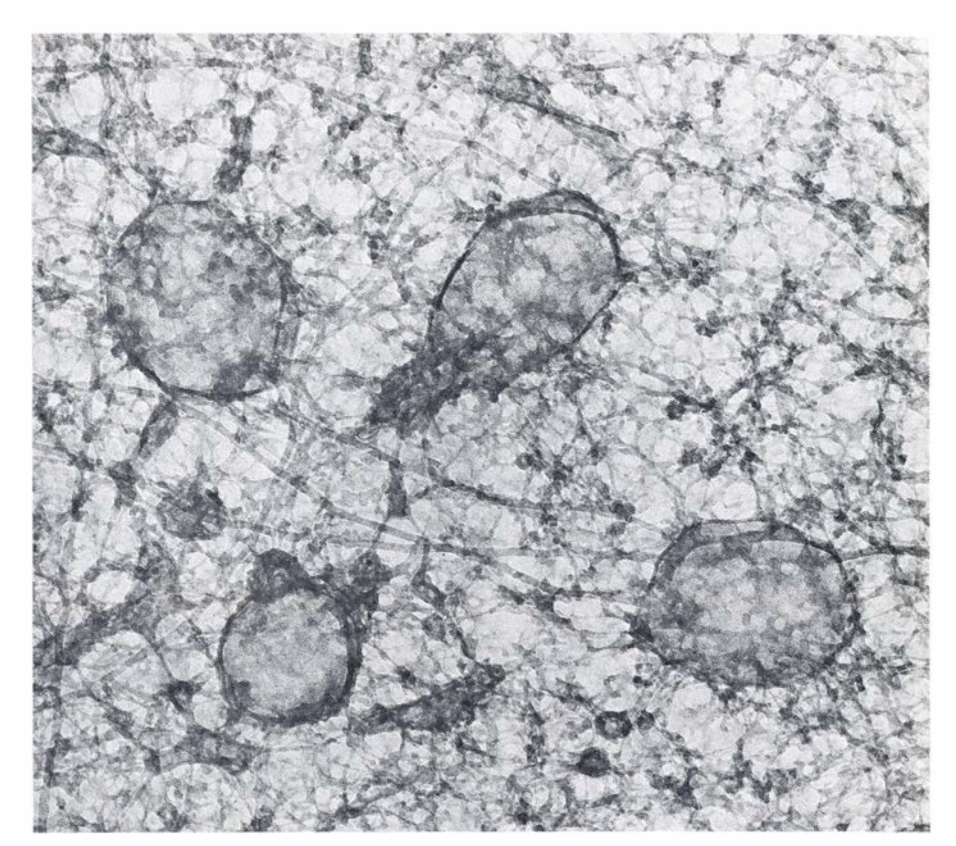
MICROGRAFIA ELECTRONICA CONVENCIONAL de un corte fino del armazón citoplasmático de una célula humana cultivada. No revela la existencia del retículo microtrabecular, aunque pueden distinguirse filamentos gruesos, además de fragmentos del retículo endoplasmático y mitocondrias. El armazón fundamental presenta numerosas fibras finas distribuidas al azar, algunas de las cuales corren por entre los microtúbulos (*estructuras largas en forma de varillas*). La resina epoxi en la que se incluyen las estructuras en el proceso de obtención de las preparaciones para microscopía electrónica vela zonas pequeñas. La micrografía, a 14.000 aumentos, fue registrada, en Boulder, por John J. Wolosewick.

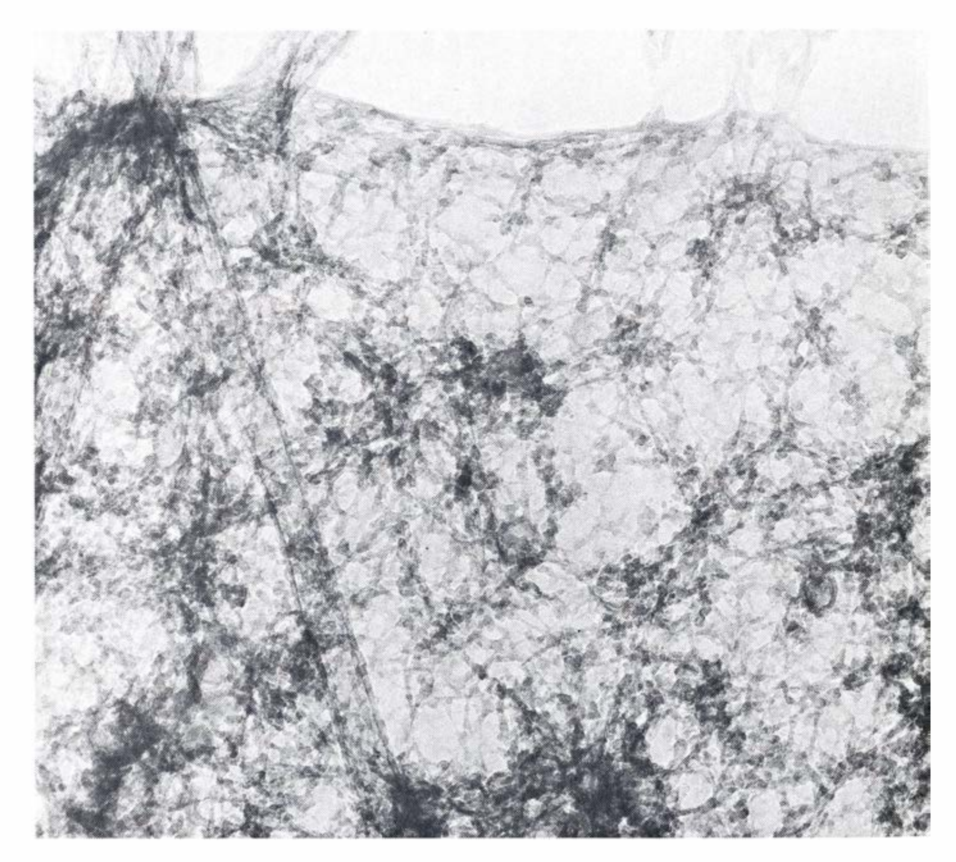
ponemos ya de datos convincentes de que diversos enzimas implicados en las vías metabólicas (por ejemplo, la vía de la glucolisis, la conversión de glucosa en piruvato) están ligados a estructuras contenidas en la sustancia citoplasmática fundamental. De ahí que tales enzimas puedan incorporarse al retículo microtrabecular, quizá según una orientación no aleatoria, de suerte que un enzima ceda su sustrato al siguiente de la vía metabólica. De esta manera, los enzimas que no están ligados a la membrana celular podrían estar coordinados espacialmente.

La fase acuosa que llena los espacios intertrabeculares está enriquecida con moléculas pequeñas: glucosa, aminoácidos, anhídrido carbónico y oxígeno. Dada la gran infiltración de la estructura reticular, el recorrido de difusión de las moléculas en la fase rica en agua es corto. Ello comporta que se acelere la velocidad a la que los sustratos van de un enzima al siguiente.

El retículo parece también desempeñar un papel importante en la diferenciación celular y en la correspondiente síntesis proteica: el proceso por el cual los polisomas (grupos de ribosomas ligados a un filamento de ARN mensajero) traducen el código genético en la secuencia específica de aminoácidos de una molécula proteica. Fundándose en electromicrografías de cortes finos se ha venido aceptando, en general, que el proceso de traducción estaba mediado por polisomas que flotaban libremente en el citoplasma. Las electromicrografías de alta aceleración revelan una morfología sorprendentemente distinta: los polisomas se hallan situados en las intersecciones del retículo microtrabecular, como arañas en una telaraña tridimensional.

Esta extraordinaria organización tiene importantes implicaciones. Significa que, tras su elaboración, las proteínas pasan a formar parte del material asociado al retículo microtrabecular. El material está entonces disponible de forma no aleatoria para la construcción controlada de microtúbulos, microfilamentos y otras estructuras celulares. Por ejemplo, los microfilamentos pueden organizarse según haces paralelos, denominados fibras de refuerzo, o según una red amorfa. Están continuamente en construcción y destrucción en diferentes partes del citoplasma, de acuerdo con la actividad que desarrolle la célula. El examen del papel del retículo microtrabecular en el conjunto de las estructuras celulares puede ser crucial para comprender cómo las células crecen y se especializan.

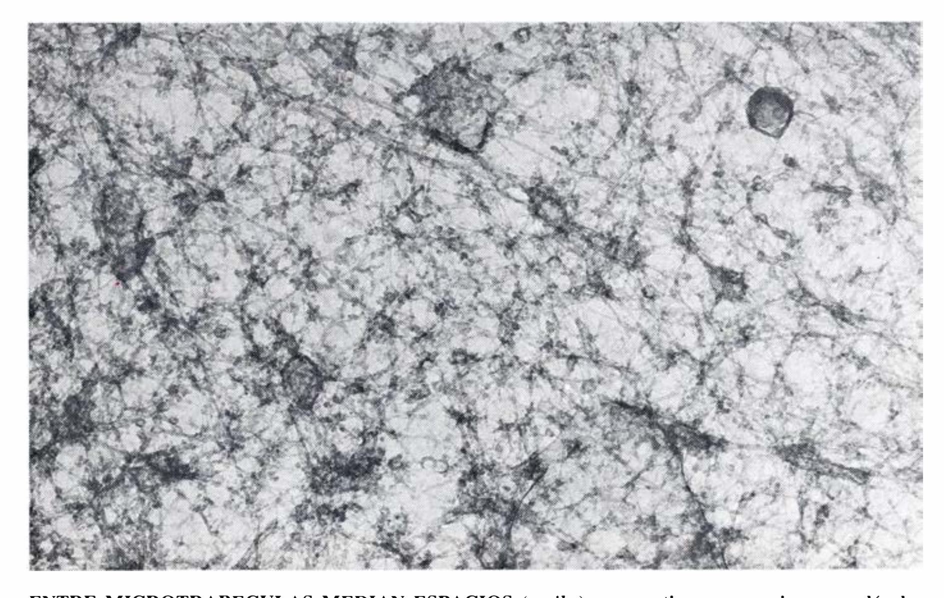




EFECTO DE LA BAJA TEMPERATURA sobre el retículo microtrabecular, resultante, quizá, de una detención de la maquinaria metabólica inducida por el frío. Las células normales enfriadas a cuatro grados Celsius durante tres horas adoptan una forma esférica. Este cambio en la configuración va acompañado de diversos cambios internos. En primer lugar, los microtúbulos se desorganizan en pocos minutos. Aparecen luego grandes espacios en el retículo microtrabecular, tal como se ve en la micrografía electrónica de alta aceleración de la parte inferior. La morfología normal se observa en la micrografía superior. Las microtrabéculas se cortan formando grandes acúmulos al ir transformándose cada una de ellas en un pequeño "muñón". Cuando las células enfriadas se reincuban a 37 grados C, los muñones se transforman de nuevo en microtrabéculas; a los treinta minutos de incubación el retículo es prácticamente normal. La sensibilidad del retículo a los cambios del entorno sugiere que es un componente celular fisiológicamente activo. Las micrografías, de Anderson, aumentan las estructuras unas 78.000 veces.







ENTRE MICROTRABECULAS MEDIAN ESPACIOS (*arriba*) **que contienen agua rica en moléculas esenciales para el metabolismo celular. (Las estructuras largas y oscuras son fragmentos de mitocondrias.) Puesto que las dimensiones de esos huecos son pequeñas, la difusión de las moléculas de un enzima a otro tarda sólo microsegundos. La existencia de espacios intertrabeculares se ha demostrado alterando la concentración del medio extracelular, forzando así la entrada o salida del agua en la célula por ósmosis. Si la concentración fuera de la célula es superior a la del citoplasma, el agua sale de la célula y los espacios intertrabeculares se reducen (*centro*). Si la concentración es inferior, el agua entra en la célula y los espacios se expanden (*abajo*). Micrografías obtenidas por Anderson; aumento: unos 35.000 diámetros.**

Dado que el retículo microtrabecular se infiltra por toda la célula, parece unir los componentes dispersos por el citoplasma –orgánulos y fibras estructurales– en una unidad estructural y funcional única, denominada citoplasto. ¿Cuál es la relación entre el retículo y la elaborada red citoesquelética formada por los tres sistemas principales de fibras? Manfred Schliwa, a la sazón en Boulder, exploró recientemente esta cuestión extrayendo células cultivadas mediante un detergente (Triton X-100) en presencia de un tampón que estabilizara los componentes citoesqueléticos. El tratamiento de las células con el detergente a lo largo de un período de sesenta segundos implicaba la extracción de prácticamente todas las estructuras envueltas por una membrana (membrana celular externa, mitocondrias, retículo endoplasmático y envoltura nuclear incluidos). Se extrajeron también casi todas las fibras microtrabeculares, lo que daba pie a pensar que sus propiedades físicas y su composición química diferían de las de los elementos citoesqueléticos estables.

El material restante era un armazón filamentoso, que se aproximaba a la forma de la célula, formado por microfilamentos, microtúbulos y filamentos intermediarios. Al microscopio electrónico de alta aceleración, la organización tridimensional de estos filamentos aparecía bien conservada y prácticamente idéntica a la de la célula intacta. Más aún, se diría que los diversos filamentos citoesqueléticos se hallaban interconectados en múltiples puntos, lo que explicaba por qué la rotura de un sistema de filamentos con inhibidores específicos podía inducir cambios en la organización de otro sistema de filamentos.

Con los elementos citoesqueléticos, los filamentos microtrabeculares parecen actuar a modo de horma de soporte de la membrana celular externa y de las extensiones altamente asimétricas de la superficie celular. Por ejemplo, los cilios y los flagelos, los apéndices en forma de látigo que propulsan a los protozoos y los espermatozoides, son complejas estructuras formadas por múltiples conjuntos de microtúbulos ligados por "radios" y "puentes". Cuando estas estructuras especializadas, incluidas en polietilenglicol, se cortan y se examinan bajo el microscopio electrónico de alta aceleración, los radios y puentes ofrecen el mismo aspecto que las microtrabéculas. De manera similar, los haces de microfilamentos de los microvilli, extensiones citoplasmáticas

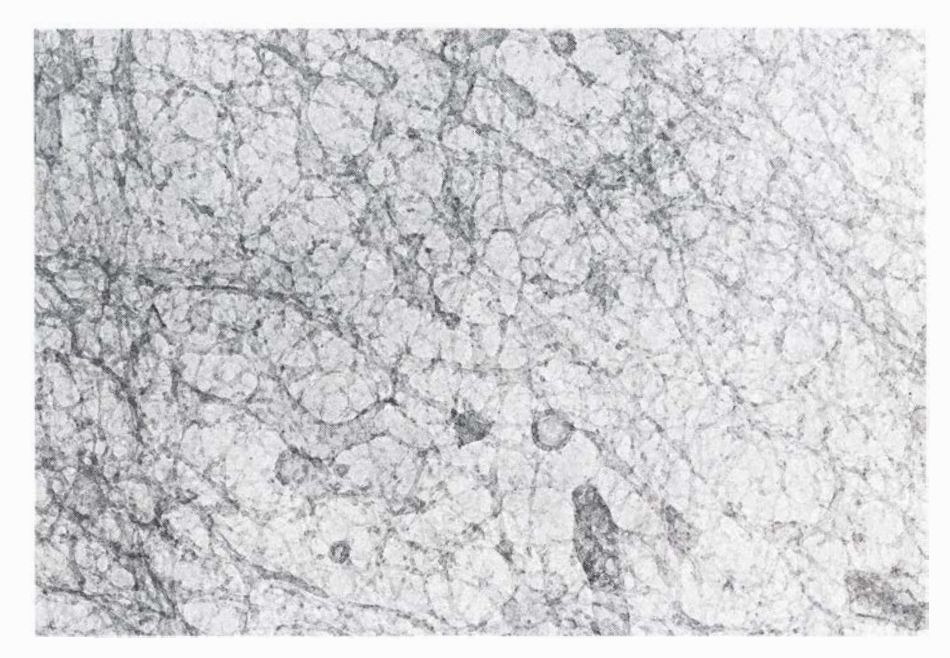
observadas en células del intestino, parecen estar asociados a microtrabéculas. Estas observaciones sugieren que la morfología interna de cilios, flagelos, microvilli y otros apéndices especializados de la célula es la misma que la del citoplasto considerado como un todo: consta de microtrabéculas y espacios intertrabeculares.

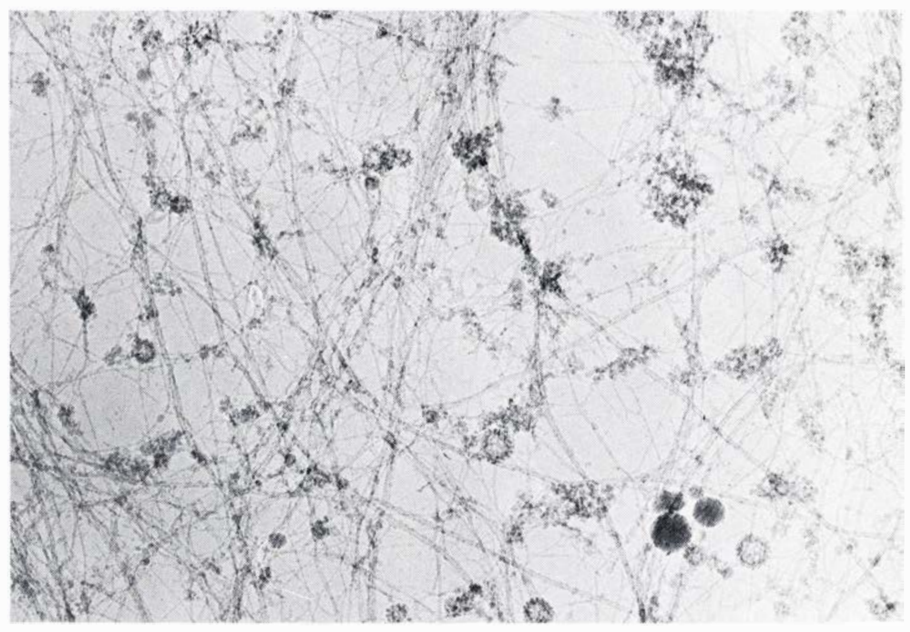
Un orgánulo citoplasmático que parece estar íntimamente asociado con el retículo microtrabecular es el centro organizador de los microtúbulos (COM), situado cerca del punto medio celular. Un COM típico observado en células animales está formado por un par de centriolos (cada uno de los cuales consiste en un conjunto de microtúbulos dispuestos según un cilindro) orientados en ángulo recto uno respecto al otro y rodeados por un complejo de proteínas satélite densas que varían en forma, número y disposición de un tipo celular a otro. Los centriolos y sus satélites sirven de puntos de iniciación para la construcción de microtúbulos. El esquema de distribución de estos elementos citoplasmáticos determina la forma de la célula.

Los COM se prolongan, sin solución de continuidad, en el retículo microtrabecular; no existe una demarcación estructural definida entre ellos. Además, la información que establece la localización de este complejo en el centro celular parece estar incorporada en el armazón fundamental. En la región citoplasmática inmediata que rodea al COM, el retículo microtrabecular se muestra extremadamente denso. Lo que explica que esta parte de la célula, conocida como centrosfera, sea muy gelatinosa. La densidad del retículo decrece, hacia la periferia celular, con la distancia respecto al COM.

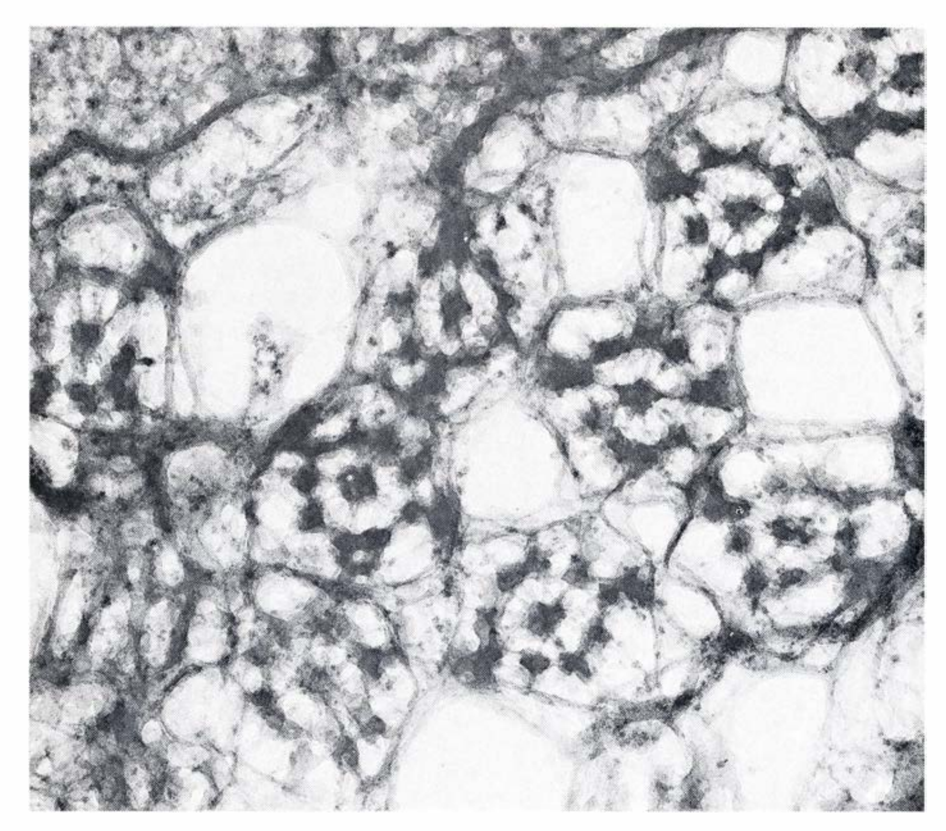
En las primeras etapas de la duplicación celular la centrosfera se divide. Las dos partes resultantes forman los polos del huso mitótico, que dirige la separación de los cromosomas replicados en el transcurso de la mitosis, o división celular. Al final de la mitosis, cada célula hija contiene una centrosfera rodeada por su propio retículo microtrabecular. Esta observación plantea una cuestión provocadora: ¿cuál es el origen de la información estructural almacenada en el retículo? ¿Está codificada en los genes o pasa, de una manera física, de una generación a la siguiente en el transcuro de la división celular durante el desarrollo embrionario? Curiosamente, cierta información experimental sugiere que esta última posibilidad puede ser la correcta.

En 1964, T. M. Sonneborn y sus colegas, de la Universidad de Indiana, realizaron elegantes experimentos sobre la herencia de estructuras celulares en el protozoo ciliado *Paramecium*. Este animal unicelular es bastante complejo, y posee, entre otras características, numerosos cilios unidos a la zona cortical celular: la parte del retículo microtrabecular situada justo por debajo de la membrana celular externa. Sonneborn y sus colaboradores alteraron, mediante microcirugía, el modelo normal de estructura de estas células, obteniendo, por ejemplo, células con dos o más citostomas o citoproctos, o zonas corticales con cilios de polaridad invertida. Los investigadores de Indiana esperaban que las alteraciones provocadas en la estructura celular normal resultaran letales o se ocuparan de su rápida corrección los genes celulares. ¡Cuál no sería su sorpresa al comprobar que las anormalidades inducidas eran heredadas por la progenie en sucesivas divisiones celulares! Los cambios se perpetuaron a través de más de 700 generaciones de células.

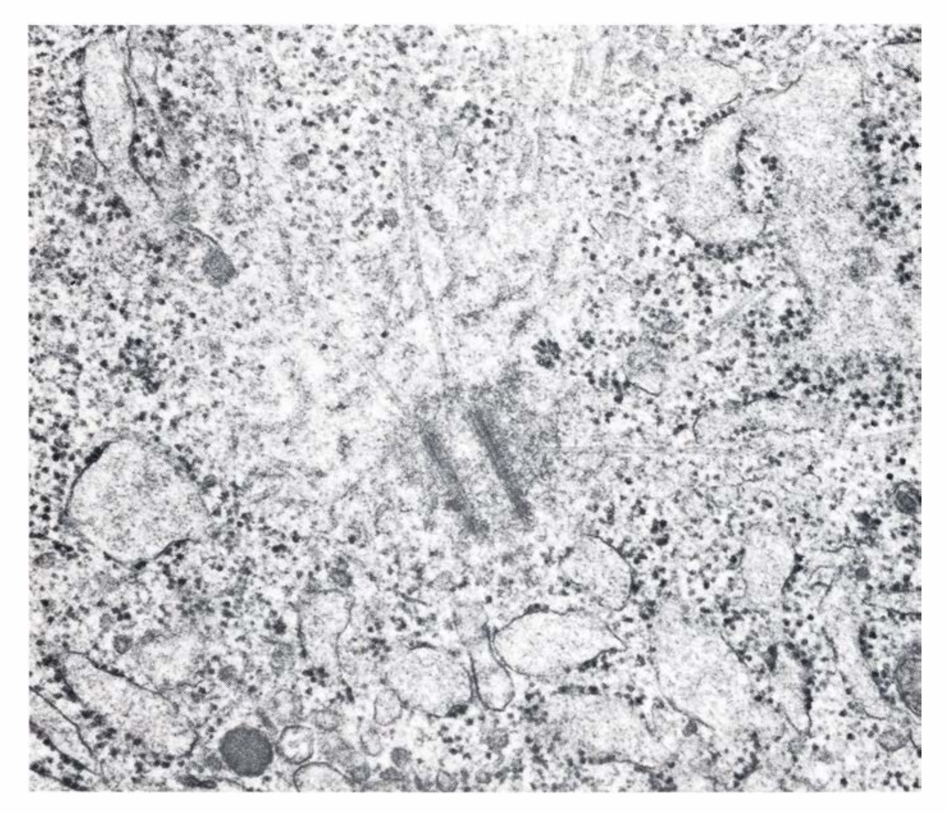




EL TRATAMIENTO CON DETERGENTE ilustra la relación entre las microtrabéculas y los filamentos del citoesqueleto (microtúbulos, microfilamentos y filamentos intermediarios). La micrografía superior muestra la célula no tratada, con el retículo microtrabecular y los orgánulos intactos. La micrografía inferior recoge una célula similar tratada con el detergente Triton X-100, que extrae el material reticular que cubre los filamentos citoesqueléticos, orgánulos y membranas. Restan los elementos citoesqueléticos, dispuestos en una red cuya forma se asemeja a la de la célula. Los microtúbulos, microfilamentos y filamentos intermediarios pueden distinguirse en razón de su grosor. Las micrografías, de unos 40.000 aumentos, las obtuvo Schliwa en el microscopio de alta aceleración de la Universidad de Boulder.



FLAGELOS DE ESPERMATOZOIDES DE RATON, vistos transversalmente en una microelectrografía de alta aceleración. Están formados por conjuntos de microtúbulos unidos por unos puentes a modo de radios, muy parecidos a las microtrabéculas. Este dato indica que las expansiones especializadas del citoplasma, como los flagelos, los cilios y los microvilli, comparten la estructura reticular que infiltra el resto de la célula. La micrografía, cuyo aumento es de 55.000 diámetros, fue tomada por Wolosewick.



CENTRIOLO, representado por las dos estructuras paralelas de la parte central, tal como aparece en una micrografía electrónica convencional de un corte fino de una célula. En una célula se distinguen, normalmente, dos centriolos, orientados en ángulo recto uno respecto al otro. Los centriolos y su material "satélite" forman un complejo que dirige la construcción y orientación de los microtúbulos. Se aprecian aquí tres microtúbulos, que arrancan del material satélite que rodea las estructuras paralelas del centriolo. El retículo adquiere mayor densidad en la vecindad del complejo. Aumento: unas 36.000 veces.

Sonneborn llegó a la conclusión de que la información de la forma se transmitía de una generación a otra a través de un mecanismo no genético: la información básica de la forma de la célula está almacenada en la estructura de la sustancia citoplasmática fundamental, que pasa a las dos células hijas durante la mitosis. Esta ordenación y disposición de la nueva estructura celular bajo la influencia de una estructura celular preexistente se denominó citotaxis. Sonneborn escribió: "Mientras los genes determinan los bloques de construcción moleculares y, mediante sus propiedades, los tipos de asociaciones moleculares que pueden darse, las asociaciones que realmente se presentan dependen, asimismo, de las que ya existen".

La idea de que la estructura citoplasmática y la motilidad estaban relacionadas partió de uno de los primeros citólogos, el francés Félix Dujardin, quien propuso, en 1835, que todas las células se componían de un material llamado sarcodio, dotado de propiedades estructurales y contráctiles. Según parece ahora, es el retículo microtrabecular el que cumple ese cometido, a modo de musculatura celular dinámica que continuamente redistribuye y reorienta los orgánulos y las fibras citoesqueléticas de acuerdo con el tipo de actividad en que esté empeñada la célula. El retículo no se mantiene, bajo ningún punto de vista, estático. Antes bien, se le supone en cambio constante, sufriendo contracciones y deformaciones locales que dan por resultado aumentos y disminuciones del tamaño de los espacios intertrabeculares.

Las células pueden efectuar diversos tipos de movimiento interno, desde el lento y continuo de los cromosomas durante la mitosis hasta el discontinuo, de retroceso, de vesículas y gránulos a lo largo de las fibras nerviosas. Los cromatóforos, o células pigmentadas, guardan un especial interés para los biólogos celulares, pues igual realizan movimientos intracelulares suaves que bruscos. La conversión de la energía química y eléctrica en trabajo mecánico para el transporte de los gránulos de pigmento en el interior de las células pigmentadas constituye, por tanto, tema importante de investigación.

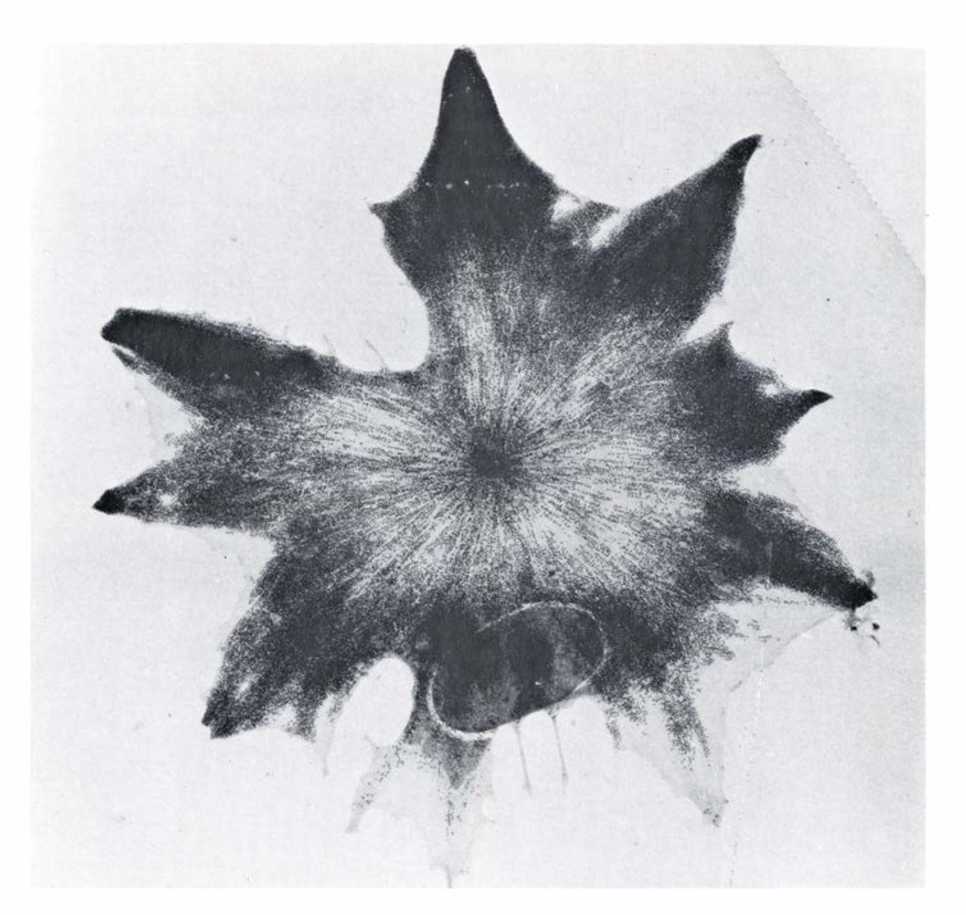
Los cromatóforos, que apreciamos en la piel de ciertos anfibios y en las escamas de muchos peces y reptiles, operan bajo control hormonal y nervioso. Les permiten oscurecer o aclarar el color de su piel de acuerdo con el de su entorno, eludiendo así a los depredadores. En los experimentos realizados por el grupo de Boulder se escogieron las

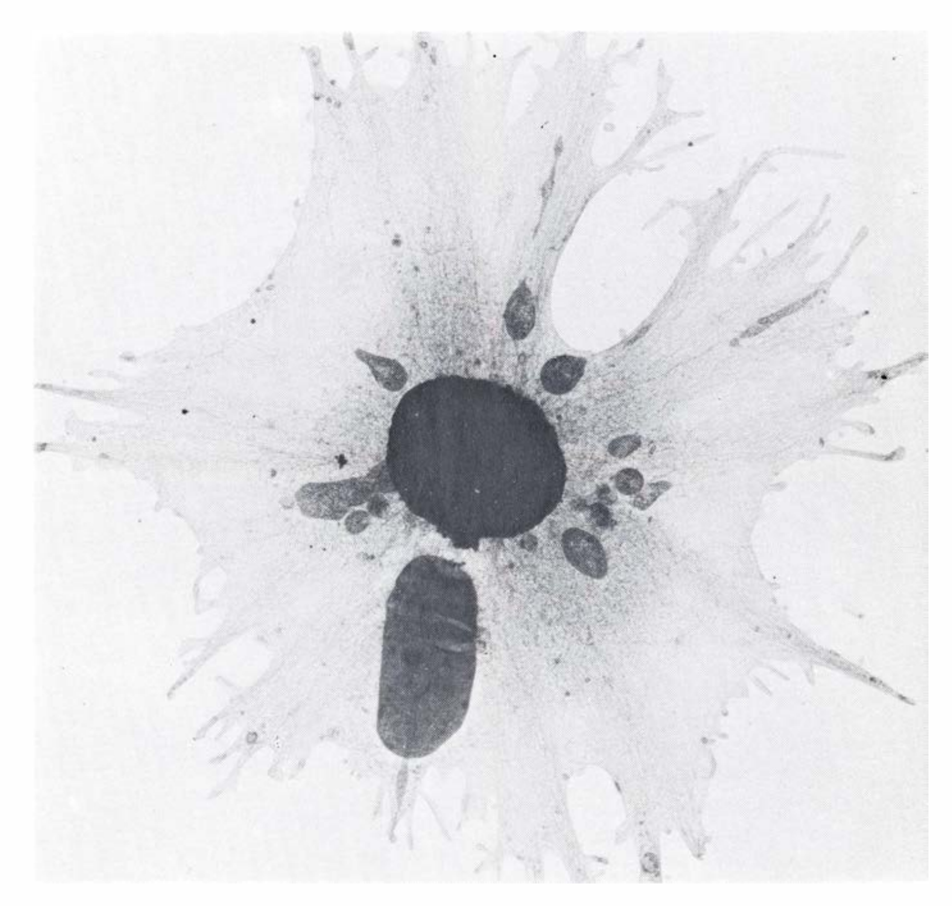
células con pigmento rojo, eritróforos, del pez *Holocentrus ascensionis.* Estas células pueden extraerse fácilmente y se mantienen aisladas en cultivo tisular. Miles de gránulos de un color rojo vivo, dispersos por todo el citoplasma del eritróforo, se agregaban en el centro de la célula en respuesta a la hormona epinefrina. Los gránulos se aceleraban hasta una velocidad de unos quince o veinte micrometros por segundo y recorrían la distancia en línea recta y sin detenerse.

La velocidad constante de los gránulos en agregación, aunque análoga a la del movimiento de los cromosomas en la división celular, supera unas 500 veces su rapidez. A los tres o cinco segundos, la agregación ha llegado a término. Los gránulos se encuentran agrupados en una masa abultada en el centro celular, rodeados por citoplasma claro, a la vez que la célula mantiene su contorno original casi circular. El núcleo celular y las mitocondrias no acompañan a los gránulos en su movimiento, sino que permanecen en el citoplasma periférico o se desplazan independientemente hacia el centro celular. Cuando la agregación se presenta simultáneamente en millones de eritróforos, el pez puede cambiar su color, de rojo vivo a blancuzco, en cuestión de segundos. La dispersión del pigmento es más lenta que la agregación; necesita de seis a diez segundos para completarse, y cada gránulo se desplaza de manera discontinua, con muchas paradas e incluso movimientos de retroceso.

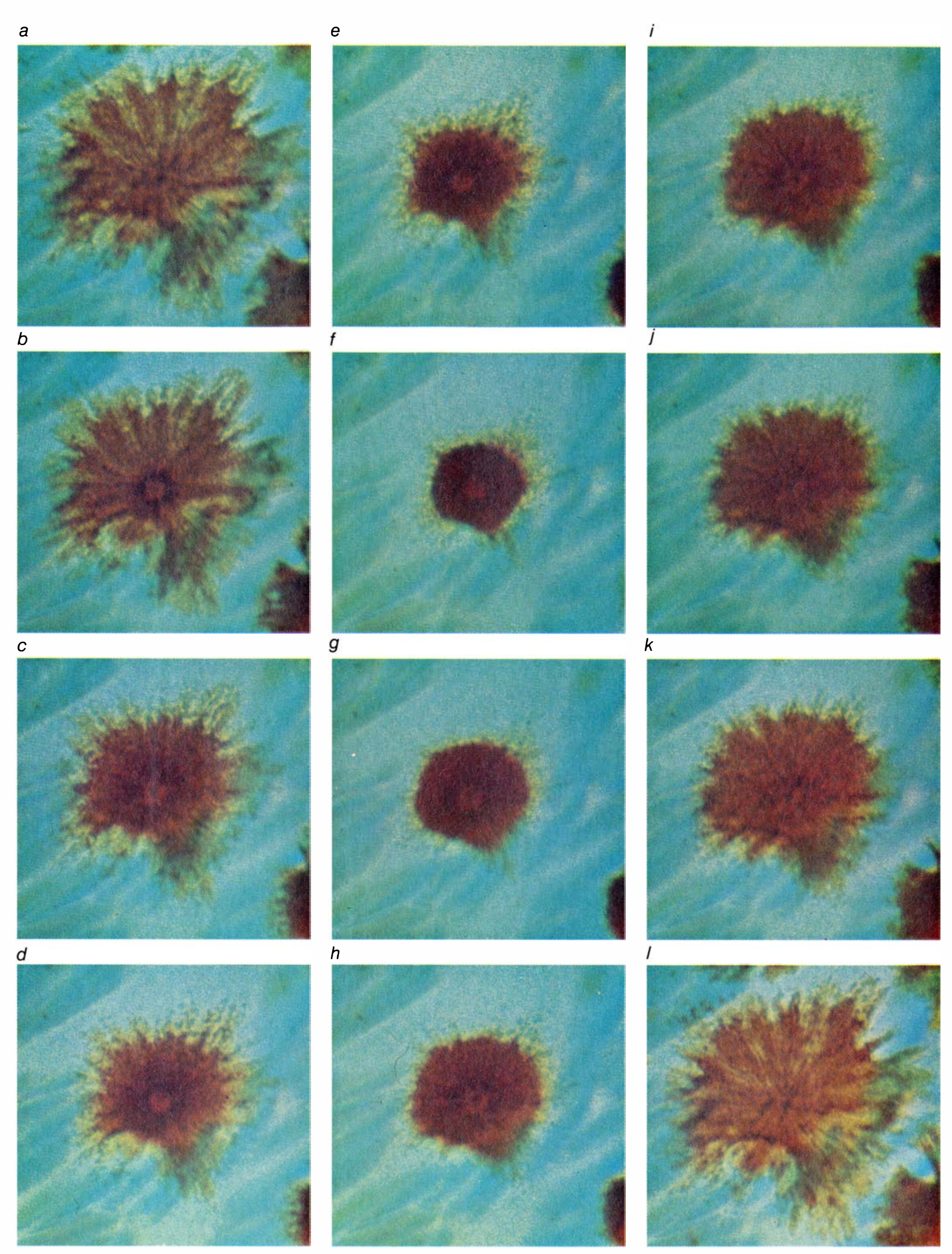
Los eritróforos agregan su pigmento en condiciones de baja energía celular, como cuando la célula muere o cuando la hemos expuesto, por largo tiempo, a la acción de un inhibidor metabólico. El proceso de agregación está, pues, impulsado por la liberación de energía potencial almacenada; en el extremo opuesto, la dispersión consume ATP y almacena energía potencial. La lentitud y discontinuidad de la dispersión derivan, seguramente, de ese carácter: hallarse en función de la energía.

Se sabía ya, por micrografía electrónica tradicional, que el eritróforo contenía una agrupación radial de microtúbulos que se extendían, cual rayos solares, desde el centro celular. En razón de esa morfología, hubo quienes, en las postrimerías de la década de los 60, propusieron que la agregación del pigmento se generaba en virtud de algún tipo de interacción por deslizamiento entre los microtúbulos y los gránulos de pigmento. Hipótesis considerada hoy poco probable, ya que apenas hay, si es que los hay, contactos íntimos entre microtúbulos y gránulos.





MOVIMIENTO DE LOS GRANULOS PIGMENTARIOS en un eritróforo, o célula pigmentaria roja, del pez *Holocentrus ascensionis*, ilustrado por dos electromicrografías de alta aceleración de células enteras. En la superior, los gránulos de pigmento se alinean según la distribución radial de microtúbulos. El núcleo de la célula (*objeto en forma de tableta*) está situado en la periferia de la célula y no parece entorpecer la alineación de los gránulos. En la célula inferior, los gránulos se agrupan en una masa densa en el centro celular. Las micrografías, de unos 3500 aumentos, las realizó Katherine J. Luby-Phelps.

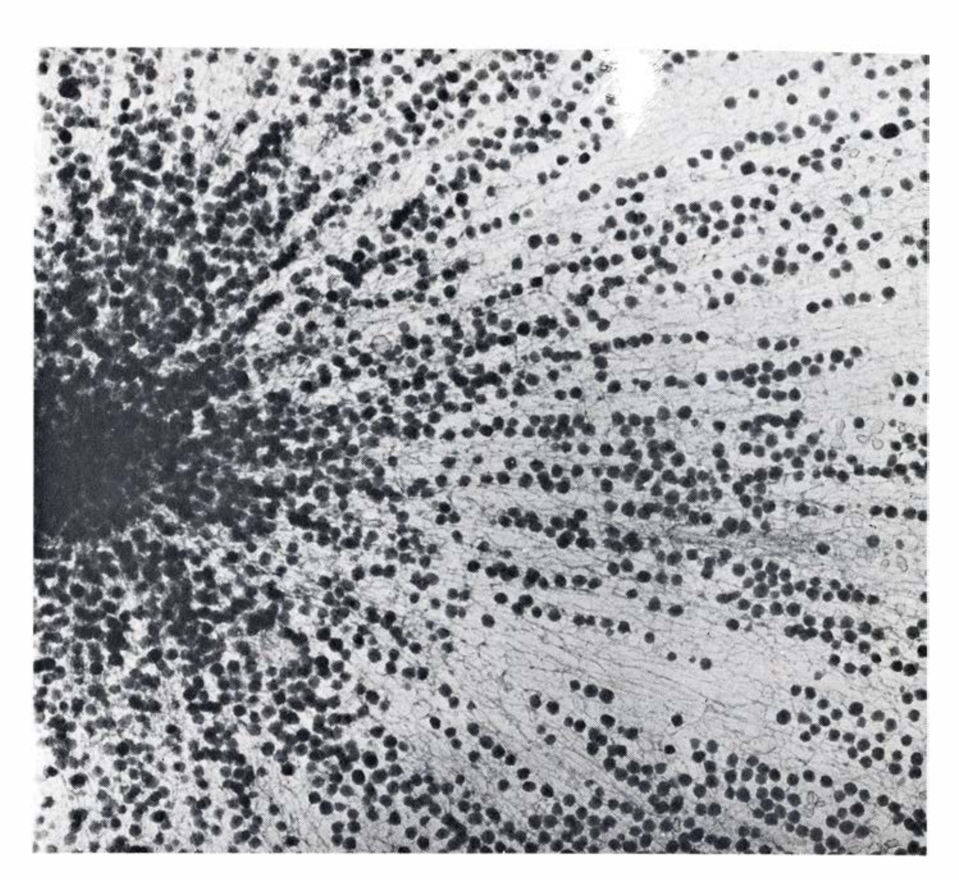


AGREGACION Y DISPERSION de los gránulos pigmentarios rojos de un eritróforo de *Holocentrus ascensionis*, según una serie de imágenes correspondiente a una película tomada con microscopio fotónico. El aumento es de 150 diámetros. La agregación simultánea del pigmento en millones de eritróforos distribuidos por todas las escamas del pez le permite cambiar su color de rojo a blancuzco en cuestión de segundos. Las imágenes *a-f* representan el proceso de agregación, que implica un movimiento rápido y continuo (velocidad uniforme) de los gránulos hacia el centro celular y que dura de tres a cinco segundos. Las imágenes *g-l* muestran el proceso de dispersión, un movimiento más lento y discontinuo (velocidad no uniforme) de regreso a la periferia celular, que se completa en unos ocho segundos, consume ATP y almacena energía potencial. La película fue tomada por Judith Fleming, en Boulder.

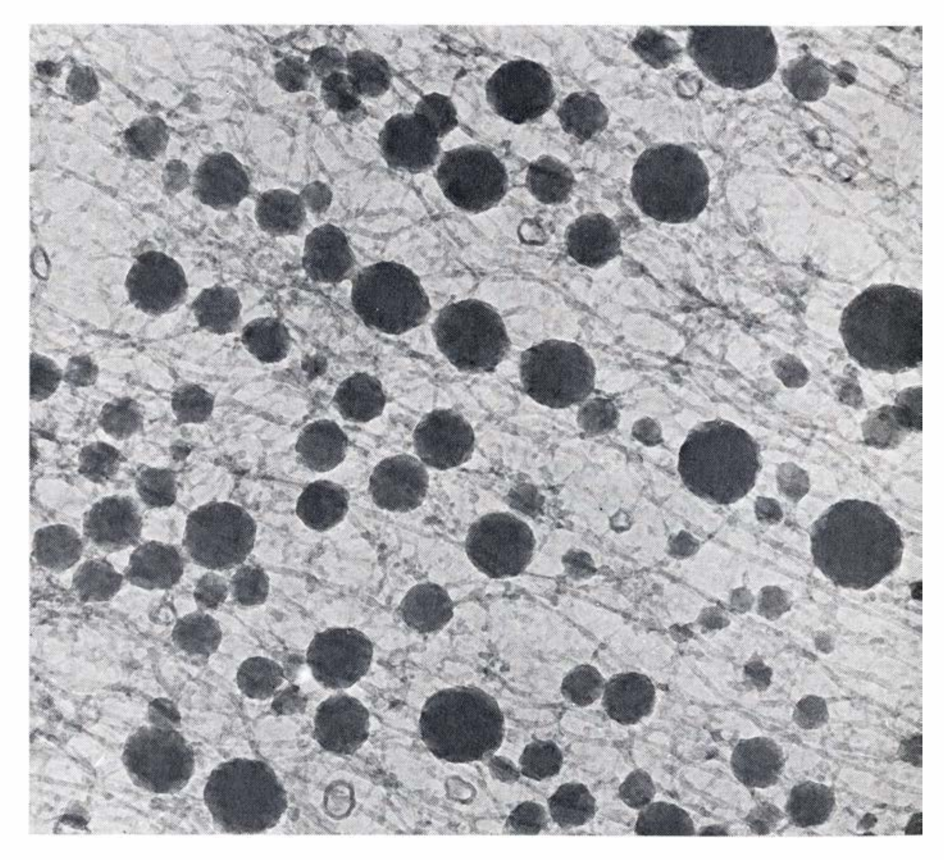
En 1977, el grupo de Boulder (al que se había unido H. Randolph Byers) se aprestó a explorar la posible conexión entre el retículo microtrabecular y la migración del pigmento en el eritróforo. Para ello examinó al microscopio electrónico de alta aceleración eritróforos aislados intactos. También en estas células hallaron una red de relleno de fibras microtrabeculares, largas y finas, que conectaban los gránulos de pigmento, los microtúbulos y las membranas celulares. Decidieron profundizar más en el estudio del mecanismo de migración pigmentaria: fijaron eritróforos en diversas fases de agregación y dispersión pigmentaria y los examinaron bajo el instrumento de alta aceleración. Descubrieron que el movimiento de los gránulos se acompañaba de notables cambios estructurales del retículo.

Durante la agregación, las microtrabéculas se acortaban y aumentaban su grosor, agrupando los gránulos y transportándolos pasivamente hacia el centro de la célula según rutas lineales definidas por la distribución radial de los microtúbulos. Al contraerse el retículo, numerosas microtrabéculas cortaban sus conexiones con los gránulos y se retraían hacia la superficie de los microtúbulos y de la zona cortical celular, para formar muñones de unos 20 nanómetros de diámetro. Una vez retirados de la periferia celular los gránulos y el material reticular asociado, la zona cortical celular superior se colapsaba sobre la inferior. En el transcurso de este último proceso, los microtúbulos dispuestos radialmente, que no se desplazaban con los gránulos, podían distinguirse sin dificultad. El retículo, por el contrario, sólo se apreciaba entonces en aquellas regiones de la célula inmediatamente adyacentes a la masa granular central y a lo largo de algunos haces de microtúbulos. En esas zonas, el retículo era muy compacto: las microtrabéculas presentaban aproximadamente la mitad de su longitud normal y aparecían en forma de rosario a intervalos regulares con muñones de 20 nanómetros.

Al iniciarse la dispersión, las microtrabéculas se alargaban, transportando los gránulos de pigmento hacia la periferia celular y reconstruyendo el retículo. Simultáneamente, los muñones que habían permanecido en la periferia de la célula se alargaban. Junto con los muñones, la disposición microtubular parecía actuar a modo de "memoria" estructural, que guiaba la reconstrucción del retículo. Ciertamente, si se desmoronaba la población de microtúbulos con colchicina, los gránulos se desorganizaban y no podían dispersarse.



DISPOSICION LINEAL DE LOS GRANULOS PIGMENTARIOS en el eritróforo dispersado, según aparece en las electromicrografías de alta aceleración. El denso amasijo de gránulos de la izquierda corresponde al centro celular. Haces de microtúbulos se proyectan, desde el centro, hacia la periferia; podemos apreciar el retículo microtrabecular de la sustancia fundamental entre los gránulos y los microtúbulos. La micrografía fue obtenida, también en Boulder, por Luby-Phelps, a unos 17.500 aumentos.



GRANULOS ESFERICOS DE PIGMENTO, que aparecen incluidos en el retículo microtrabecular en esta micrografía de alta aceleración obtenida por Luby-Phelps. Los gránulos están recubiertos por material reticular y algunos de ellos se ofrecen, con toda nitidez, enlazados por microtrabéculas. Los microtúbulos cruzan diagonalmente el campo desde la parte superior izquierda hasta la inferior derecha.

Los trabajos realizados en Boulder por Katherine J. Luby-Phelps han reforzado la hipótesis de que el retículo media la migración del pigmento. Estudió cómo repercutía en la migración del pigmento el empleo de inhibidores metabólicos que agotasen el suministro de ATP a los eritróforos. Las células tratadas presentaban un retículo de aspecto desarticulado, caracterizado por una gran proporción de microtrabéculas incompletas o con su extremo libre. Sólo podían dispersar parcialmente el pigmento. En ausencia de ATP, los filamentos microtrabeculares parecían incapaces de alargarse y reestructurar el retículo. Según se desprende, pues, el gasto de energía química (ATP) permite el almacenamiento de energía potencial en el retículo durante la dispersión, energía que se transforma en cinética durante la agregación. Las idas y venidas de los gránulos pigmentarios en el eritróforo en reposo pueden resultar de una continua lucha entre la tendencia entrópica del retículo a contraerse y su capacidad de expandirse en respuesta a un suministro energético, como si se tratara de una goma elástica sometida a estiramiento.

El estado de contracción del retículo quizás esté controlado por los iones calcio, pues se sabe que éstos regulan la contracción muscular y diversas actividades motiles en células no musculares. Luby-Phelps ha observado que un aumento en el nivel de iones calcio libres en el eritróforo dispara la contracción de las microtrabéculas y conduce a la agregación del pigmento. La necesidad de ATP durante la dispersión puede deberse, al menos en parte, a la puesta en marcha de un mecanismo de captación que elimine los iones calcio libres de los espacios intertrabeculares, de modo que las microtrabéculas puedan volver a su estado elongado.

La fuente más probable de iones calcio libres es un sistema de canales tubulares encerrado en membranas, que se desliza por entre los gránulos de pigmento del eritróforo. Un sistema poligonal de estos túbulos se extiende por toda la célula, en contacto con el retículo y, en algunos lugares, con la zona cortical celular superior e inferior. Un sistema de este tipo podría suministrar el calcio a todos los puntos de la célula en respuesta a un impulso nervioso. (El eritróforo tiene muchas conexiones nerviosas.) A este respecto, el sistema tubular del eritróforo podría funcionar de manera análoga al retículo sarcoplasmático del músculo, que responde a un impulso nervioso descargando calcio, lo que dispara la contracción muscular.

La explicación completa de cómo el retículo microtrabecular se expande y contrae vendrá cuando hayamos alcanzado un conocimiento pormenorizado de su composición molecular y estructura. Obviamente, esta organización bioquímica debe permitir un comportamiento elástico y, al mismo tiempo, promocionar la formación de múltiples interconexiones, con el propio retículo y con los microtúbulos, microfilamentos, filamentos intermediarios y orgánulos.

Si se someten las células a la acción de un detergente, y las proteínas extraídas se separan mediante electroforesis bidimensional en gel, se distinguen más de cien proteínas. Ello indica que el retículo es un complejo polímero de muchas proteínas. De momento se han identificado ya algunas: la actina y la miosina (las principales proteínas contráctiles de las células musculares, presentes también en células no musculares), la tubulina (el elemento de construcción de los microtúbulos) y dos proteínas específicas asociadas a los microtúbulos, entre otras. Aunque gran parte de la estructura de la matriz puede representar una asociación laxa de proteínas fácilmente extraíbles, hay pruebas de la existencia de una verdadera espina dorsal filamentosa en el retículo. Muchas microtrabéculas podrían tener un núcleo de actina filamentosa envainada en otras proteínas.

Otro campo de investigación prometedor, aunque todavía en estado embrionario, es el relativo a las diferencias que ostenta la estructura del retículo microtrabecular según se trate de células normales o de células cancerosas. El retículo de las células malignas es más denso que el de las células normales y presenta signos de una falta de organización normal, particularmente en la distribución de los polisomas.

En resumen: el armazón citoplasmático fundamental, durante largo tiempo considerado una solución homogénea y rica en proteínas, parece entrañar una estructura y un comportamiento muy intrincados. El retículo microtrabecular organiza los diversos componentes celulares según una unidad funcional (el citoplasto) y media el transporte regulado y dirigido dentro de la célula. Cada vez parece más claro que el control de la forma y el movimiento celular dependen del funcionamiento integrado de los microtúbulos, microfilamentos, centros organizadores de los microtúbulos y del retículo microtrabecular. El estudio de la estructura y fisiología celulares adquiere ahora nuevas perspectivas.

Las erupciones del monte Santa Elena

El ciclo de la actividad actual del volcán forma parte de otro ciclo mayor, de 4500 años de duración. Sus violentas erupciones del año pasado fueron predichas por los vulcanólogos que se encontraban en la zona

Robert Decker y Barbara Decker

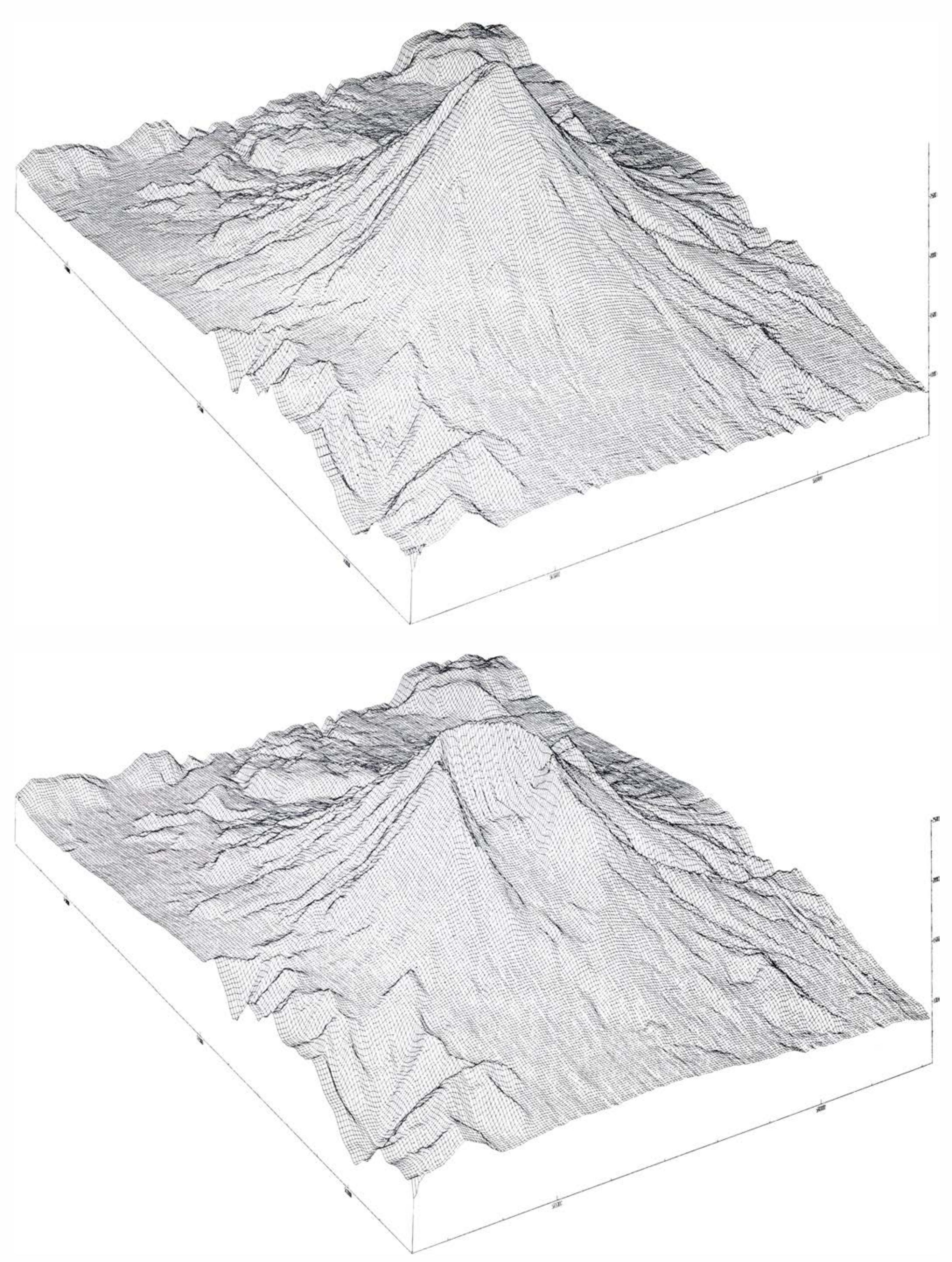
Una de las erupciones volcánicas más estrechamente vigilada y de la que se han tomado más fotografías, además de haberse obtenido abundante información directa, fue la violenta erupción del monte Santa Elena, ocurrida el 18 de mayo de 1980. También fue la primera erupción volcánica registrada en los 48 estados contiguos de la Unión desde que tuvieron lugar, entre 1914 y 1917, las erupciones mucho menos importantes de Lassen Peak. La del Santa Elena desplazó 2,7 kilómetros cúbicos de rocas volcánicas (entre los que cabe contar 0,5 kilómetros cúbicos de magma nuevo, es decir, rocas en estado líquido), que devastaron una zona de más de 500 kilómetros cuadrados, produciendo una de las mayores avalanchas que registra la historia. En los últimos cien años, la explosión del monte Santa Elena sólo ha sido superada, en magnitud, por las erupciones del Santa María (Guatemala, 1902), del Krakatoa (Indonesia, 1883) y del Katmai (Alaska, 1912) que emitieron, respectivamente, unos 5, 6 y 12 kilómetros cúbicos de magma (volúmenes reducidos, si atendemos a la densidad de la roca solidificada).

La misma erupción del Katmai resulta insignificante, comparada con otras erupciones más antiguas cuyas enormes deposiciones de ceniza y roca, que alcanzaron un espesor comprendido entre decenas y centenares de metros, cubrieron miles de kilómetros cuadrados. Así ocurrió, por ejemplo, en Japón, Nueva Zelanda, América Central, en el oeste estadounidense y en muchas otras regiones volcánicas del mundo. El volumen de los materiales expulsados por esas erupciones descomunales osciló entre 100 y más de 1000 kilómetros cúbicos. ¿Fue el vulcanismo prehistórico, en punto a intensidad, superior a la intensidad máxima que hoy en día cabe imaginar, o es acaso que la experiencia humana sobre el particular abarca un lapso de tiempo demasiado corto? La mayoría de geólogos se inclinan por lo segundo. Las erupciones del último siglo fueron, probablemente, pequeñas muestras de la energía volcánica que la tierra encierra aún en su interior.

Desde que se produjo la erupción del monte Santa Elena se ha intentado cuantificar la energía en ella liberada, comparándola con la energía desprendida en las explosiones nucleares. La comparación es útil, pero se presta a ciertos equívocos. No sólo se trata de una fuente de energía completamente distinta, sino que difiere también el ritmo de su liberación, es decir, la potencia. La energía térmica y mecánica liberadas en el monte Santa Elena el 18 de mayo de 1980 se cifró en torno a los $1{,}7 \times 10^{18}$ joule. Puesto que una explosión nuclear de un megatón libera unos $4{,}2 \times 10^{15}$ joule, la erupción del Santa Elena fue equivalente a una explosión de 400 megatones, es decir, unas ocho veces superior a la mayor explosión nuclear jamás producida. La comparación puede inducir a engaño, porque casi toda la energía de la explosión nuclear se transforma en energía mecánica y térmica en un momento, de modo que la potencia casi instantánea en watt de una explosión nuclear tiene esencialmente el mismo valor que la energía en joule (un joule por segundo equivale a un watt). En la erupción del Santa Elena, los $1{,}7 \times 10^{18}$ joule se disiparon en nueve horas; ello viene a equivaler a una potencia media de unos 5×10^{13} watt ($1{,}7 \times 10^{18}$ dividido por 32.400 segundos). La producción continua de energía por parte del volcán podría parangonarse, pues, con la detonación de una serie de 27.000 bombas del tipo de las que cayeron sobre Hiroshima, es decir, una cada segundo durante nueve horas. Utilizando otra comparación se puede afirmar que la energía generada por el Santa Elena el 18 de mayo fue del orden de unas 100 veces superior a la capacidad de producción de todas las centrales eléctricas de los Estados Unidos.

El monte Santa Elena ha entrado en erupción en 20 ocasiones, como mínimo, a lo largo de los últimos 4500 años, justificando así la denominación que recibió de los indios del noroeste americano: Loowit (Dama de Fuego). Antes de las erupciones de 1980, el Santa Elena había entrado en actividad, por última vez, entre 1831 y 1857. Se trata de uno de los 15 volcanes principales de la cordillera de las Cascadas que, desde el Lassen Peak (California), se dirige, hacia el norte, hasta el monte Garibaldi (Columbia Británica). Dicha cordillera forma parte del "círculo de fuego", conjunto de cordilleras volcánicas que rodean, sin apenas solución de continuidad, el océano Pacífico. En las zonas de subducción esas cordilleras cabalgan las placas tectónicas en movimiento de la cuenca del Pacífico, que se hunden, por debajo de las primeras, por el oeste, norte y este. La subducción de las placas del Pacífico produce violentos terremotos y aporta las rocas fundidas y la presión necesarias para activar los volcanes del círculo de fuego. Ciñéndonos a los tiempos históricos, se cifra en unos 400 el número de los volcanes que se han mostrado activos en los bordes de la cuenca del Pacífico.

La importancia de la actividad volcánica parece estar regida, en parte, por la velocidad de las placas que sufren subducción y se hunden bajo las placas marginales. En Indonesia y Japón, donde la velocidad de la subducción es

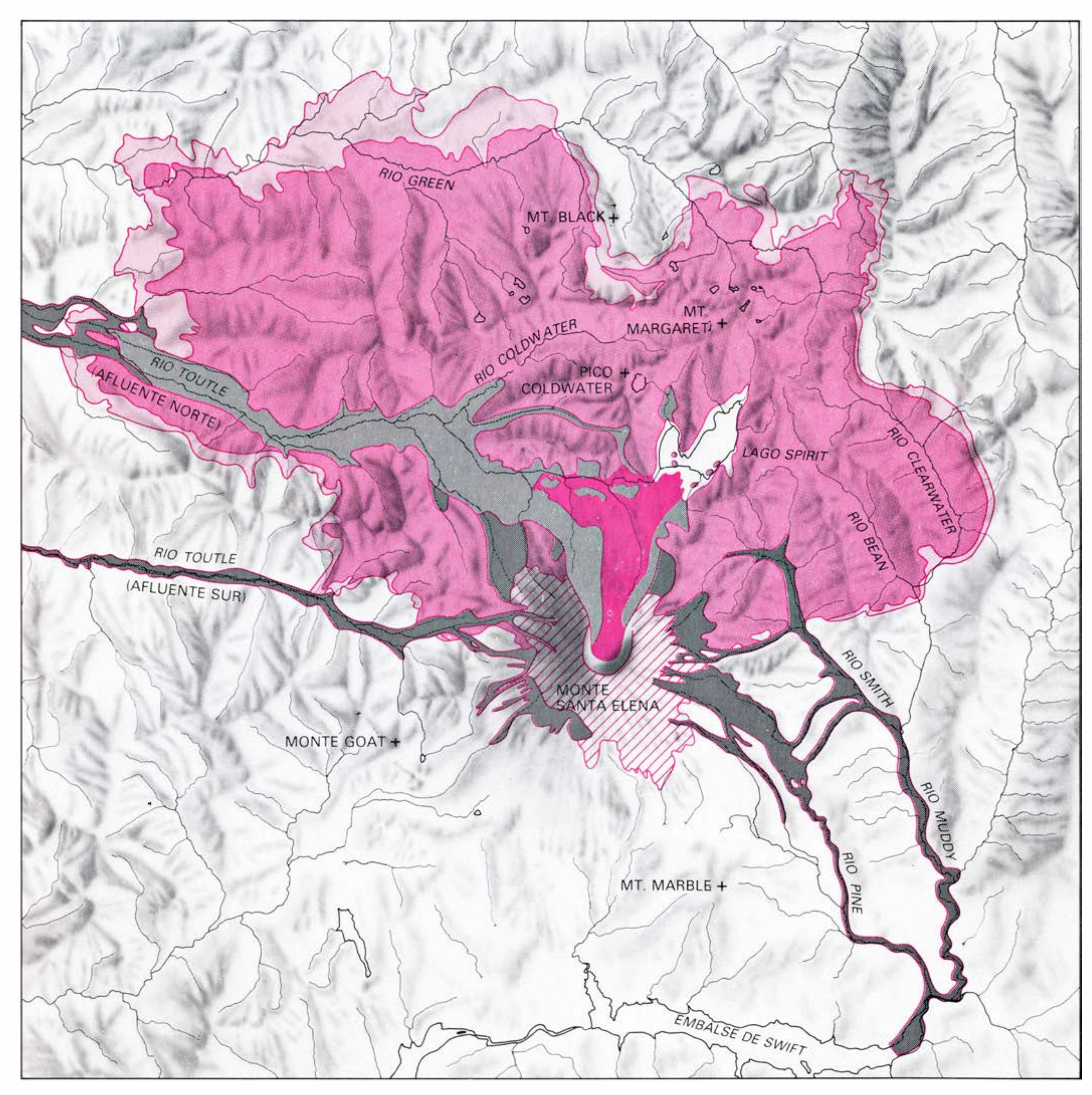


MODELOS GRAFICOS, realizados mediante ordenador, del monte Santa Elena antes de la erupción explosiva del 18 de mayo de 1980 (*arriba*) y después de ella (*abajo*). Se aprecian los efectos de la erupción que eliminaron los últimos 400 metros de la cumbre de la montaña, dejando un cráter de 750 metros de profundidad. La nube de cenizas alcanzó ese día más de 20 km de altura. Estos bloques-diagrama cartografían el accidente geográfico desde el nordeste. Fueron realizados por el programa Digital Elevation Model en el Western Mapping Center del Servicio Geológico de los Estados Unidos.

de seis o siete centímetros por año, se asiste, como mínimo, a una erupción volcánica por año. En la cordillera de las Cascadas, la baja frecuencia con que se producen las erupciones parece estar relacionada con la velocidad de convergencia, pequeña, de la placa norteamericana y de la de Juan de Fuca, situadas inmediatamente al oeste; dicha velocidad de convergencia es de dos o tres centímetros por año.

A principios de la década de 1960, Dwight Crandell y Donal Mullineaux, del Servicio Geológico de Estados Unidos, intentaron valorar los riesgos que comportaban los volcanes apagados, aunque activos en potencia, de la cordillera de las Cascadas. Comenzaron su programa con un estudio de los depósitos volcánicos existentes en las cercanías del monte Rainier, situado 75 kilómetros al nornordeste del Santa Elena. El Rainier había entrado en erupción, en alguna ocasión, entre 1820 y 1854. Guiados por las dataciones de erupciones anteriores, hechas con carbono-14, y siguiendo la opinión de los geólogos de que lo sucedido antes puede repetirse de nuevo, Crandell y Mullineaux predijeron, de forma general, los riesgos potenciales de cada uno de los volcanes que estudiaron.

El informe de Crandell y Mullineaux sobre el monte Santa Elena apareció en 1978. Llegaron a la conclusión de que tenía unos antecedentes pésimos: durante los últimos 4500 años había sido el más activo y explosivo de cuantos existían en los 48 estados contiguos de la Unión. En este período, entre sus materiales eruptivos se encontraban domos de lava demasiado viscosa para salir de sus cráteres, grandes lluvias de cenizas volcánicas con gran cantidad de

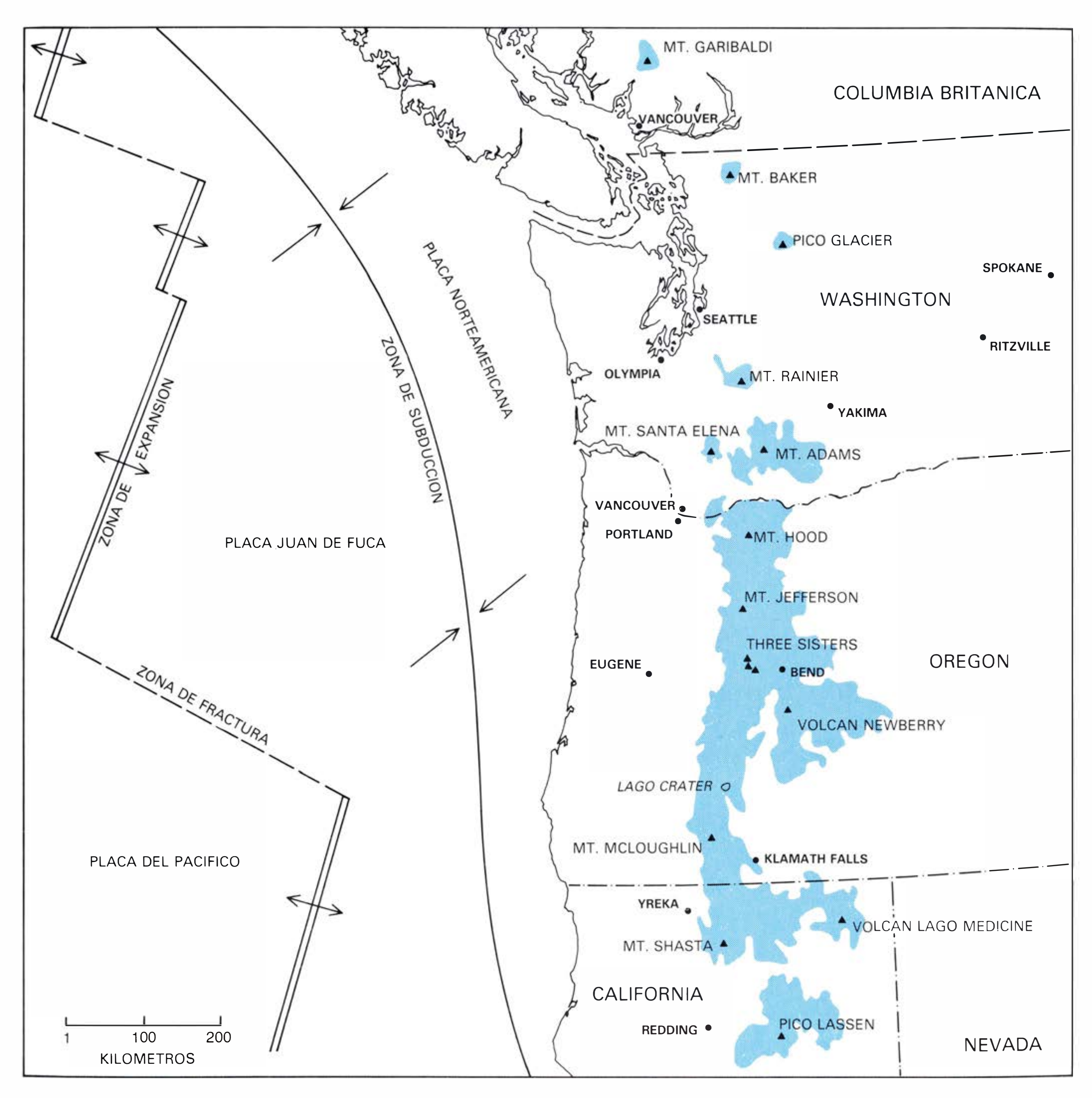


MAPA GEOLOGICO de la región devastada el 18 de mayo. La zona de color intenso representa los depósitos piroclásticos. La zona rayada corresponde a las caídas de ceniza y, la de color intermedio, a la extensión donde los árboles fueron derribados. En la zona de color claro los árboles permanecieron en pie, pero sus hojas murieron por el calor. El color gris claro señala los depósitos formados por las avalanchas y, el gris oscuro, las coladas de barro.

trozos de pumita, coladas de piroclastos (fragmentos a elevada temperatura cuarteados y fluidizados por la actividad eruptiva), coladas de lava y coladas masivas de barro que se desplazaban por los valles fluviales que irradiaban del volcán. El intervalo medio entre los períodos eruptivos era de 225 años.

Basados en su propia investigación, Crandell y Mullineaux avanzaron: "en el futuro, el monte Santa Elena entrará probablemente en erupción intermitentemente y de forma violenta, tal como lo ha venido haciendo en la historia geológica más reciente. Estas futuras erupciones atentarán contra la salud y la vida humanas, la propiedad, la agricultura y la prosperidad económica, en general, de una zona extensa... Las características propias del comportamiento del volcán sugieren que el intervalo actual de quietud no tendrá una duración larga, del orden del milenio, y lo más probable es que se produzca una erupción en los 100 próximos años, quizás incluso antes de finalizar el siglo".

Muchas erupciones volcánicas van precedidas de enjambres de terremotos de pequeña intensidad. Si bien no todos los volcanes de la cordillera de las Cascadas disponían de una vigilancia sísmica, el Santa Elena sí la tenía afortunadamente. La Universidad de Washington había instalado, en el flanco oeste del volcán, un sismógrafo que estaba conectado con Seattle por telemetría. El 20 de marzo de 1980, a las 3 horas 47 minutos de la tarde (hora del Pacífico), se produjo bajo la montaña un terremoto de magnitud 4 en la escala Richter. Cuando a este acontecimiento tan poco frecuente le siguieron un número cada vez mayor de terremotos locales, se vio claramente que se estaba desencadenando un importante



MAPA TECTONICO, donde se muestra la relación existente entre la placa norteamericana y las de Juan de Fuca y del Pacífico, situadas ambas al oeste. En la zona de subducción, la placa de Juan de Fuca se hunde bajo la placa de Norteamérica, originando los volcanes de la cordillera de las Cascadas (*triangulitos negros*). Las partes coloreadas indican depósitos volcánicos de edad inferior a dos millones de años. Datos recogidos por el Servicio Geológico.



FOTOGRAFIAS TOMADAS ANTES Y DESPUES de los fenómenos volcánicos del monte Santa Elena, donde se observa la extensión del nuevo cráter. Realizó la superior uno de los autores (Robert Decker) en junio de 1970, cuando la cumbre, que se distingue desde el nornordeste, tenía una altura de 2950 metros, estando su base situada a unos 1000 metros de altura. La fotografía inferior fue tomada por Ray Foster, de los laboratorios Sandia, en julio de 1980. El cráter, visto desde el norte, tiene dos kilómetros de diámetro. La altitud de su borde superior está comprendida entre 2400 y 2550 metros y la de su base entre 1800 y 1900. Las coladas de piroclastos recubren la mayor parte de la zona del primer plano de la fotografía. La devastación es total.

enjambre de seísmos. Con objeto de mejorar los registros y localizar los temblores, se instalaron sismógrafos adicionales.

El 25 de marzo, la energía sísmica desprendida por el enjambre alcanzó su ritmo máximo al registrarse 47 terremotos de magnitud 3, o superior, en un período de 12 horas. Los terremotos se producían a poca profundidad bajo la zona de la vertiente norte de la montaña. Las primeras explosiones pequeñas de vapor de agua llegaron dos días más tarde, empezando a las 12 horas 36 minutos de la noche y dando origen a un nuevo cráter, de unos 70 metros de diámetro, en la zona de la cumbre que estaba cubierta de nieve y hielo. Se formaron también, en la nieve y el hielo, grandes grietas de dirección este-oeste, señal de que en la zona de la cumbre se estaba produciendo el hundimiento de un bloque por falla. El 29 de marzo se abrió un segundo cráter. De noche y desde el aire se podían observar llamas azuladas, debidas, seguramente, a la combustión de sulfhídrico. El 30 de marzo se contabilizaron noventa y tres pequeñas erupciones de vapor de agua y ceniza.

El primero de abril, los sismógrafos registraron el primer temblor volcánico, vibración del suelo más o menos continua que se observa en muchos volcanes activos. Se desconoce la causa exacta que origina un temblor de este tipo, si bien se cree que refleja el movimiento del magma o el ruido producido por la liberación de gas previamente disuelto en el mismo.

Para entonces, la nueva erupción volcánica, la primera que se producía en la cordillera de las Cascadas desde que el Lassen Peak se apagara en 1917, ya había atraído la curiosidad de la gente. Nolan Lewis, director de los Servicios de Emergencia del distrito de Cowlitz (estado de Washington), donde se encuentra el monte Santa Elena, informó que "el domingo (30 de marzo) la buena visibilidad permitía distinguir las carreteras que suben hasta la montaña; parecían tan concurridas como las calles de Seattle a las horas punta". Cuando el peligro de que se produjeran erupciones violentas se fue haciendo cada vez más evidente, el gobernador, Dixy Lee Ray, ordenó reforzar el bloqueo de las carreteras que conducen hasta el monte.

Las pequeñas erupciones de vapor de agua y ceniza continuaron produciéndose, algunas de ellas en forma de explosiones aisladas, mientras que otras se manifestaban como surtidores rítmicos de varias horas de duración. Las columnas de vapor y ceniza se elevaban hasta tres kilómetros por encima de la cumbre. El cráter se agrandó hasta formar una sola hondonada de 500 metros de largo por 300 de ancho y 200 de profundidad. Las cenizas estaban compuestas por fragmentos de roca volcánica antigua y las emisiones de gas, constituidas por pequeñas cantidades de anhídrido carbónico, anhídrido sulfuroso, sulfhídrico y ácido clorhídrico, además de grandes cantidades de vapor de agua. La razón de que ocurrieran erupciones explosivas de poca importancia obedecía, seguramente, a que el agua subterránea existente en las partes altas del cono volcánico, calentada por encima de su punto de ebullición, se transformaba súbitamente en vapor, en un proceso parecido a lo que acontece en un geyser, pero con la energía suficiente para incorporar partículas de ceniza y producir un cráter de explosión. La energía desprendida por todas las explosiones de vapor de agua hasta el 18 de mayo se estima en 10^{14} joule.

Mientras persistía la elevada actividad sísmica (unos 50 terremotos de magnitud 3, o superior, por día) se produjo otro hecho de carácter amenazador. El 27 de mayo, varios observadores habían notado ya que el hundimiento por falla del bloque de la cumbre de la montaña parecía ir acompañado de una prominencia o levantamiento del elevado flanco norte del cono volcánico. La prominencia continuaba aumentando a principios de abril, manifestándose gracias a una red cada vez mayor de grandes grietas que afectaba al manto de hielo y nieve. Los mapas fotogramétricos realizados por el Servicio Geológico de los Estados Unidos, a partir de fotografías aéreas tomadas el 12 de abril, demostraron que la prominencia tenía un diámetro de casi dos kilómetros y que ya había experimentado un levantamiento del orden de los 100 metros.

De las campañas sobre el terreno llevadas a cabo a finales de abril y principios de mayo se dedujo que la prominencia continuaba aumentando hacia el norte, de forma más o menos horizontal, a razón de unos 1,5 metros por día. La zona de deformación rápida se encontraba directamente por encima del centro de la zona sísmica, situada dos kilómetros más abajo. La opinión general de los observadores que allí se hallaban era que la persistencia de los enjambres sísmicos y la importante deformación de la topografía constituían una prueba clara de que el magma se iba inyectando a poca profundidad bajo el volcán. Esperaban que se produjera una erupción importante o una avalancha en la cara norte en expansión. El único problema consistía en saber cuánto tardaría y cuál sería su violencia.

Robert Christiansen, investigador encargado del programa de vigilancia del Servicio Geológico de los Estados Unidos, analizó la actividad secuencial histórica de otros volcanes, especialmente la del volcán de Lassen Peak, el único de la cordillera de las Cascadas con una erupción bien observada. Hizo también lo propio con el de Bezymianny, en la península de Kamchatka (costa pacífica de la Unión Soviética), que sufrió una violenta explosión en 1956. Christiansen concluyó que la evolución más probable que seguiría el Santa Elena se asemejaría a la erupción del Lassen, de 1915. No podía excluirse, sin embargo, una erupción de las características del Bezymianny. Cabía, en una tercera posibilidad, que cesara toda actividad sin que se produjera una erupción de importancia.

Entretanto, el 7 de mayo, después de unas dos semanas de poca actividad visible, volvieron las pequeñas explosiones de vapor de agua y ceniza. Aunque la actividad sísmica no había remitido y la prominencia había ido aumentando de forma constante, la falta de una actividad visible espectacular indujo a los residentes en esta región a poner en entredicho el mandato gubernativo que mantenía cerrada la zona aledaña. Durante los días 15, 16 y 17 de mayo continuaron produciéndose los terremotos y creció la prominencia, pero no había indicios de vapor de agua o de gases.

A las 7 en punto de la mañana, hora del Pacífico, del 18 de mayo, Dorothy y Keith Stoffel, geólogos de Washington, subieron en un avión ligero en el aeropuerto de Yakima, cercano al monte Santa Elena, y despegaron para realizar su primer vuelo de reconocimiento. Hicieron varias pasadas alrededor y por encima del volcán. No apreciaron actividad alguna, por más que la mañana era soleada y la visibilidad, buena. A las 8 horas 32 minutos, un terremoto de magnitud 5,1, localizado debajo del flanco norte, sacudió la montaña. En ese instante preciso, los Stoffel se encontraban encima de la cumbre, observando desde una altura de 400 metros. Se fijaron en cómo empezaban a caer fragmentos de hielo de las abruptas la-

deras del cráter. Quince segundos más tarde, se convirtieron en los testigos más próximos del inicio de una enorme erupción volcánica desencadenada por uno de los mayores deslizamientos de terreno que registra la historia.

"Todo el flanco norte del cráter de la cumbre comenzó a convulsionarse instantáneamente cual masa gigantesca", explicaría luego Dorothy Stoffel. "La masa se contraía y agitaba sin avanzar lateralmente. El flanco septentrional entero de la cumbre empezó luego a desplazarse hacia el norte, a lo largo de un plano de deslizamiento subterráneo."

Unos segundos más tarde se produjo una gran explosión. Es curioso que los Stoffel no la sintieran ni oyeran, a pesar de encontrarse exactamente al este de la cumbre. Desde su posición, la nube de la explosión inicial parecía adquirir, lateralmente hacia el norte, la forma de hongo y descender luego por las laderas. No había tiempo que perder, si querían mantenerse con vida. A pesar de lanzarse en picado a toda marcha, para ganar velocidad, la nube grisácea en expansión les alcanzó; por fin escaparon de ella virando hacia el sur. Detrás suyo, un hervidero gigantesco de nubes de ceniza iba ganando altura, impulsadas hacia el norte y el noroeste. Hacia el este, las nubes, que no paraban de crecer, adquirían formas de hongo ondulante iluminadas por rayos.



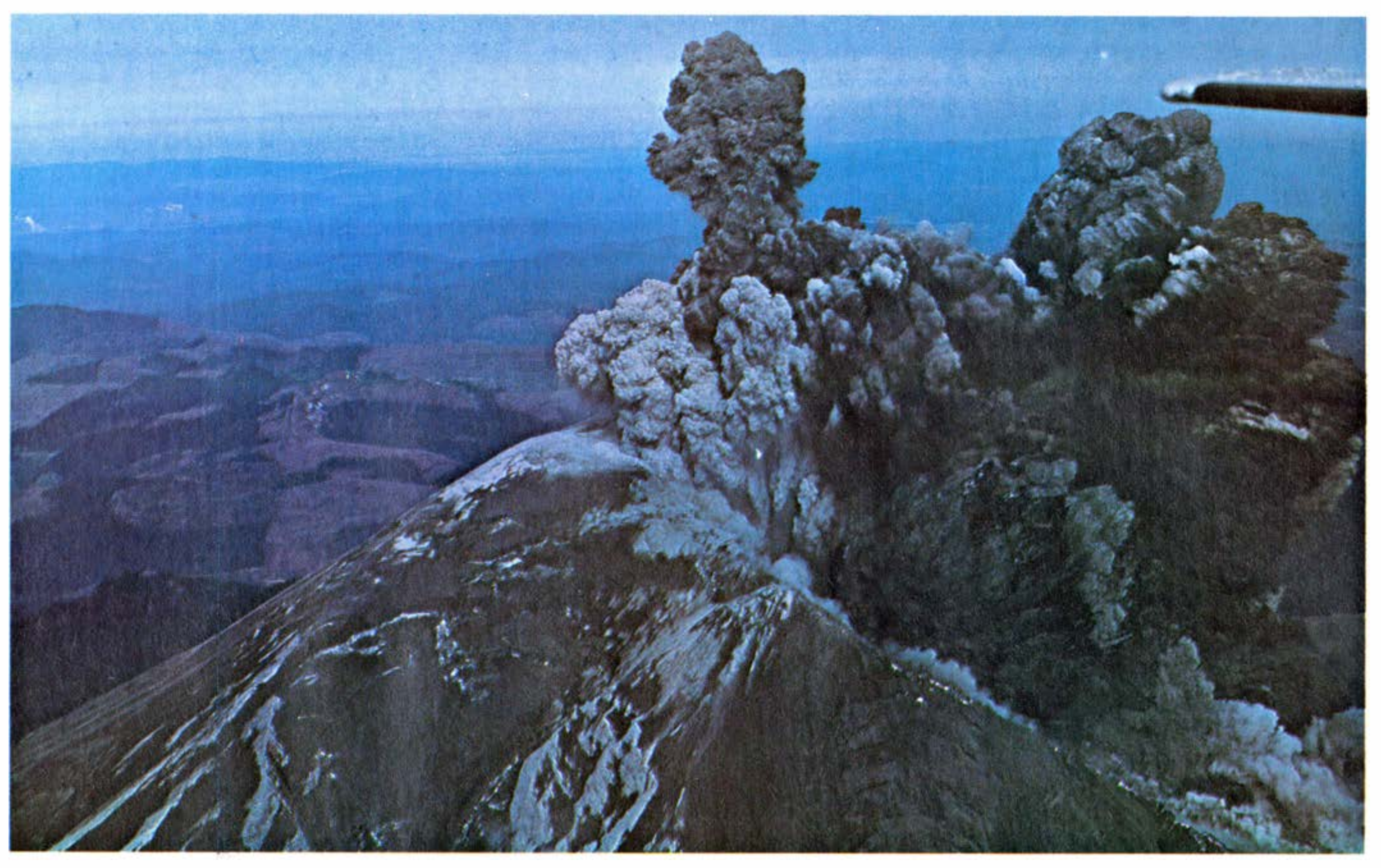
VISTAS VERTICALES del monte Santa Elena, tomadas antes y después de la erupción, el 1 de mayo de 1980 (*izquierda*) y el 19 de junio de 1980 (*derecha*), por un avión U-2 de la NASA. La cumbre se encuentra en la parte inferior derecha. La película utilizada es sensible al infrarrojo y los positivos

En la avalancha que se había formado, y de la que habían sido testigos desde el aire, más de dos kilómetros cúbicos de rocas cuarteadas y de hielo de los neveros se precipitaron en el lago Spirit y en el afluente norte del río Toutle. Al fluidificarse por el vapor de agua en explosión, la avalancha sufrió una rápida aceleración, alcanzando velocidades del orden de 250 kilómetros por hora. Un lóbulo de esta masa gigantesca se encajó en el brazo occidental del lago Spirit, siguiendo hacia el norte por el valle que se continúa en esa dirección. Otro lóbulo adyacente barrió otro valle con tal ímpetu que sobrepasó una cresta de 360 metros de altura que lo limita por el norte. El grueso de las rocas fluidificadas descendió como en un embudo por el valle del río Toutle, formando un depósito de topografía irregular, accidentada por numerosos y diminutos montículos, de 21 kilómetros de largo, de uno a dos de ancho y con un espesor de hasta 150 metros. La energía gravitatoria de la avalancha fue de unos 5×10^{16} joule (el equivalente aproximado a 12 megatones).

Como la gran avalancha de hielo y roca rebajó súbitamente la presión en el interior del cono volcánico, el agua subterránea sobrecalentada se convirtió inmediatamente en vapor. Al mismo tiempo, se produjo la explosión de los gases disueltos en la masa magmática situada a poca profundidad, de re-



aparecen aquí en un color convencional. Las zonas en rojo corresponden al verde de la vegetación, formada en gran parte por abetos rojos. Las zonas grisáceas de la foto de la derecha fueron devastadas por la erupción del 18 de mayo. En esta foto, y sobre la cumbre, una nube de humo sale del cráter.



MOMENTO DE LA ERUPCION del 18 de mayo (a los 20 segundos de su inicio) registrado por Keith y Dorothy Stoffel, geólogos que sobrevolaban la montaña en aquellos momentos. El comienzo de la avalancha precedió a la erupción. Los Stoffel pudieron escapar, pese a ser atrapados por la nube.

ciente intrusión en el núcleo superior de la montaña. La explosión del vapor de agua y la del magma, junto con la gigantesca avalancha, se combinaron para originar una explosión lateral constituida por nubes de vapor de agua (con trozos de roca en suspensión) densas y a elevada temperatura (hasta 300 grados C), lanzadas hacia el norte, desde el flanco abierto de la montaña, a velocidades comprendidas entre 100 y 400 kilómetros por hora. La explosión de vapor de agua y su carga fluidificada de fragmentos de rocas volcánicas devastaron 550 kilómetros cuadrados de terreno montañoso al noroeste, norte y nordeste del monte Santa Elena. Las nubes negruzcas que barrieron el suelo rodaron por encima de cuatro valles importantes, y de sus correspondientes crestas divisorias, alcanzando zonas distantes hasta 28 kilómetros.

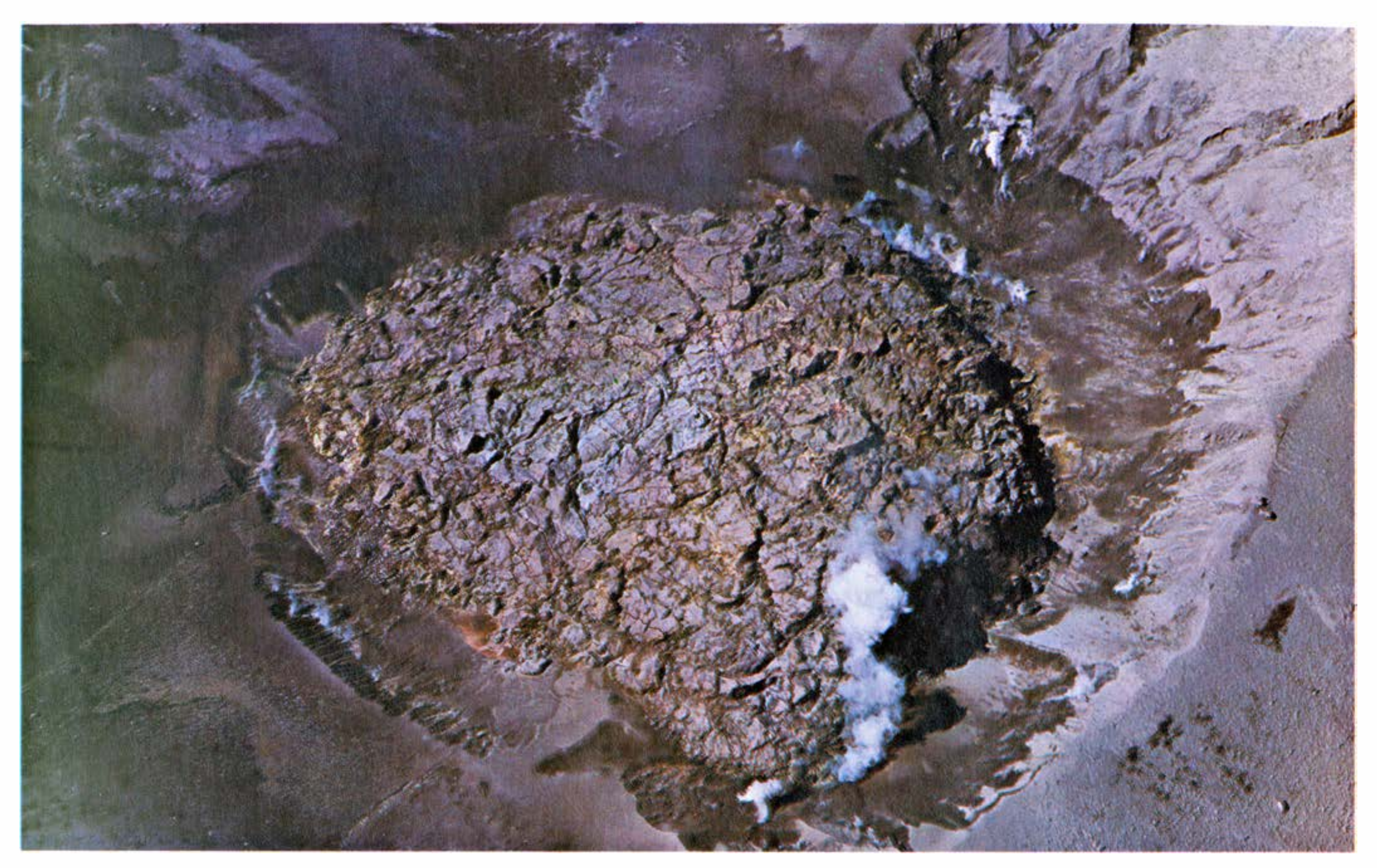
La destrucción fue total. En los primeros kilómetros fueron levantados de cuajo y barridos, por la explosión arrasadora de la nube, árboles enteros de uno a dos metros de diámetro. Más allá de esta zona se produjo otra, de 10 a 15 kilómetros de ancho, donde el derribo de árboles fue general, pudiéndose observar cómo saltaban en astillas los abetos rojos de primera calidad que allí había. En los límites más periféricos de la zona de destrucción, los árboles se mantenían aún en pie, pero sus troncos, calcinados, se perdieron sin remisión.

La primera impresión que producía la observación de la zona catastrófica era de que una especie de conmoción, o gran onda de choque, había derribado los árboles en una dirección radial a partir de la cumbre donde se había originado la explosión. Impresión que no podía mantenerse, empero, con las señales allí existentes. Los supervivientes que se encontraban en los límites de la zona devastada sólo oyeron una explosión moderadamente fuerte, un estruendo, dos o tres minutos antes de que la nube negra, con sus vientos huracanados y a elevada temperatura, descendiera sobre ellos. La velocidad del frente de las nubes de explosión cargadas de vapor de agua era muy inferior a la velocidad del sonido. De cerca, los árboles derribados presentaban una disposición en forma de remolinos de turbulencia y de líneas de corriente curvadas.

En la vecindad de la región devastada, los árboles se tumbaron hacia las pendientes descendentes del valle, incluso cuando esto implicaba que lo hacían en dirección a la fuente que originó las impetuosas nubes.

Al parecer, la gravedad activó la masa densa y fluidificada cuando disminuyó la energía de la explosión inicial de vapor de agua. Luego, al menguar los vientos turbulentos internos, las cenizas y los trozos de roca que estaban en suspensión en las densas nubes se sedimentaron sobre la superficie topográfica, formando depósitos que disminuyen de espesor a medida que aumenta su distancia a la fuente de origen, pudiendo ésta variar entre un metro y un centímetro. Fragmentos angulares de hasta varias decenas de centímetros de diámetro, tanto de rocas volcánicas antiguas como recientes, a elevada temperatura, fueron transportadas en las nubes de explosión hasta 10 y 15 kilómetros. Hasta siete kilómetros de distancia por el noroeste y norte, y 18 por el nordeste, los árboles quedaron chamuscados y se ennegrecieron por el lado encarado a la zona de explosión.

El carácter fluidizado de las masas de fragmentos transportados por la onda explosiva se hace mucho más patente en los depósitos de las vertientes empinadas. Allí, después de la deposición inicial, las coladas secundarias originaron depósitos de relleno, de decenas de



DOMO DE LAVA formado por el cráter entre el 13 y el 20 de junio. Tenía 300 metros de anchura por 65 de alto y fue destruido posteriormente por la erupción del 22 de julio. Maurice y Katia Krafft, científicos del Centro Vulcanológico de Cernay (Francia), son los autores de este documento fotográfico.

metros de espesor, en las cubetas y fondos de valle. El volumen total de los depósitos de explosión alcanzó unos 0,18 kilómetros cúbicos. De esta cantidad, unos 0,06 kilómetros cúbicos eran de origen magmático, formados por rocas volcánicas recién constituidas. La energía calorífica liberada por esta componente magmática se cifró en 2×10^{17} joule.

Otra fuente de energía, que probablemente influyó en la explosión del vapor de agua, fue la derivada de las aguas subterráneas sobrecalentadas del volcán. Admitiendo una porosidad del 15 por ciento para un volumen de dos a tres kilómetros cúbicos, empapados de agua a una temperatura media de 175 °C, ello supone una adición de energía equivalente a 10^{17} joule. El vapor de agua de la nube de explosión se originó, verosímilmente, de dos maneras. El calor de las aguas subterráneas que salieron al exterior de forma repentina habría producido $4,4 \times 10^{10}$ kilos de vapor y el calor adicional de los 0,06 kilómetros cúbicos de fragmentos magmáticos habría convertido otros $8,8 \times 10^{10}$ kilos de agua en vapor. El total de los $1,3 \times 10^{11}$ kilos de vapor de agua ocuparían un volumen de 220 kilómetros cúbicos a 100 grados C y a la presión atmosférica. La masa de agua subterránea sobrecalentada que se supone existía en el interior del volcán ($3,75 \times 10^{11}$ kilos) venía a triplicar la cantidad necesaria para suministrar el volumen calculado de vapor de agua. Esta energía y los valores del volumen de vapor son sólo estimaciones aproximadas, pero dan resultados razonables si los consideramos referidos a la zona de 550 kilómetros cuadrados de extensión devastada por las nubes de la explosión lateral que asolaron el terreno. El depósito final de la explosión fue una capa de ceniza húmeda, con un espesor máximo de seis centímetros, que contenía, en muchos puntos, las conocidas bolas de barro, del tamaño de un guisante, que los vulcanólogos denominan lapillis de acreción. Estas acreciones se originaron alrededor de núcleos formados por las gotas de lluvia condensadas a partir de las nubes de vapor de agua.

Hacia las 9 de la mañana del 18 de mayo había terminado lo peor de la erupción. Pero la columna vertical prosiguió rugiendo, alcanzando alturas superiores a 20 kilómetros durante la mayor parte del día, hasta que empezó a disminuir a las 5 horas 30 minutos de la tarde. La fuente originaria de esta columna de gases y ceniza, en explosión casi continua y que se elevaba rápidamente, yacía en la masa magmática, en efervescencia y a poca profundidad, que se iba engullendo, progresivamente, a profundidades mayores. La abrasión producida por la correspondiente expulsión hacia arriba continuaba agrandando el cráter en forma de herradura que la avalancha y las explosiones laterales habían formado en un principio.

Durante todo el día soplaron a gran altura vientos de componente nordeste y, a media mañana, la ceniza empezó a caer sobre las ciudades de la zona central del estado de Washington. En la ciudad de Yakima, a 150 kilómetros de distancia, la primera caída de cenizas formó una capa, de coloraciones claras y oscuras, finamente entremezcladas, constituida por fragmentos, del tamaño de los granos de arena, de rocas oscuras y de cristales de feldespato de colores más claros. Sobre esta capa se depositó otra de mayor espesor, formada por partículas del tamaño de las del limo, constituidas por vidrio volcánico. Treinta kilómetros al norte de Yakima, la caída de cenizas formó una capa de unos 20 milímetros de grosor. Hacia el

este, la ceniza de grano fino alcanzó un espesor aún mayor, llegando a más de 70 milímetros en las proximidades de Ritzville, que se encuentra a 330 kilómetros del monte Santa Elena. En este último lugar, la textura de la ceniza parecía polvo de talco.

En los aledaños de Spokane (estado de Washington) y a 430 kilómetros al nordeste del volcán, las cenizas alcanzaron sólo un espesor de cinco milímetros, pero hacia las 3 de la tarde la visibilidad se redujo a tres metros y la oscuridad era casi absoluta. Cerca del mediodía del día 19 cayeron sobre Denver trazas de ceniza. Sólo tres días tardó la nube en atravesar los Estados Unidos. El peso de la ceniza caída contabilizada equivalía a 0,15 kilómetros cúbicos de magma y representó una liberación de energía térmica de 5×10^{17} joule. Los estudios de las pequeñas cantidades de ceniza que cayeron fuera de la zona donde se midió su espesor, y la existencia de aerosoles que se mantienen aún en la estratosfera, formados por partículas muy finas de ceniza y ácido sulfúrico, señalan una dispersión adicional de magma equivalente a un volumen de 0,1 kilómetros cúbicos.

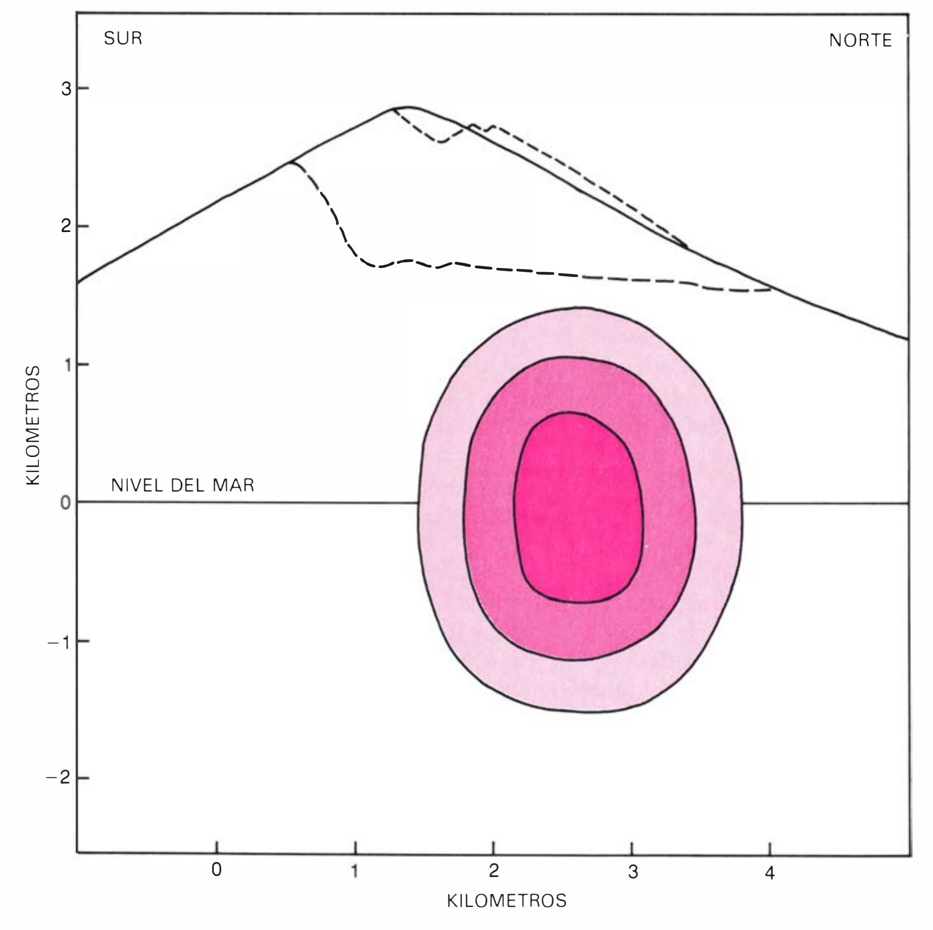
A partir de la experiencia adquirida en otras erupciones que han lanzado a la estratosfera polvo y aerosoles, puede afirmarse que esas partículas tan finas tardarán un año o dos en caer. Queda por dilucidar el efecto que ello producirá, suponiendo que produzca alguno, en el clima de la tierra.

Las inundaciones y coladas de barro constituyeron otro aspecto importante de la erupción. Las coladas estaban formadas por una mezcla líquida acuosa de cenizas volcánicas y finas partículas de roca que presentaban la consistencia del cemento húmedo. La cubierta de cenizas cercana a la montaña y las rocas cuarteadas de la avalancha suministraron el material sólido, mientras que el agua adicional procedía probablemente de diversas fuentes: de la fusión de hielo y nieve, del agua del lago Spirit y del afluente septentrional del río Toutle desplazada por el sedimento de la avalancha, del agua procedente de la ruptura del sistema hidrotermal que no se convirtió instantáneamente en vapor y del vapor condensado.

El primer máximo de las coladas de barro en el afluente meridional del río Toutle, cerca del lago Silver, se produjo a las 10 horas 50 minutos de la mañana del 18 de mayo. Sobrepasó en 30 centímetros el nivel máximo de las inundaciones registradas a lo largo de la historia. La colada de barro de mayores proporciones se produjo en el afluente septentrional del Toutle, haciéndose máxima cerca de las 7 de la tarde y destruyendo la estación de aforo cercana al lago Silver. La altura de las señales dejadas por el agua y el barro indicaban que había sobrepasado, en nueve metros, los niveles de las inundaciones de que se tengan noticia.

Aguas abajo del río Toutle, los sedimentos fangosos taponaron los cauces del río Cowlitz y produjeron una disminución muy notable de la profundidad en la fracción navegable del Columbia. Se depositaron unos 0,1 kilómetros cúbicos de barro. Aproximadamente el mismo volumen de agua intervino en la movilización de las coladas de barro.

Algo después de la avalancha inicial y de la erupción explosiva de vapor de agua, empezaron a descender rápidamente, por la vertiente septentrional del monte Santa Elena, las coladas de piroclastos, constituidas por cenizas finas y bloques de pumita, que salieron al exterior a través de las zonas de ruptura producidas en el cráter recién formado. Estas emulsiones fluidificadas de rocas a elevada temperatura y de fragmentos de vidrio mezclados con gases volcánicos sobrecalentados salieron del cráter por debajo de la nube ascendente, ya que tenían mayor densidad que esta última. Una serie de coladas sucesivas descendieron por la vertiente septentrional a velocidades de hasta 100 kilómetros por hora, recubriendo el primer deslizamiento de terreno y los depósitos de explosión, hasta alcanzar el borde meridional del lago Spirit. Cuando estas coladas de piroclastos a elevada temperatura (300-370 ºC) entraron en contacto con el agua, desencadenaron explosiones secundarias, lanzando vapor de agua y nubes de ceniza hasta dos kilómetros de altura. Las coladas de ceniza y de pumita se prolongaron hasta el atardecer del 18 de mayo, con un volumen total de 0,25 kilómetros cúbicos y una energía térmica de $3,3 \times 10^{17}$ joule.



CORTES TOPOGRAFICOS correspondientes al monte Santa Elena a lo largo de los meses siguientes: agosto de 1979 (*línea continua de la parte superior*), 1 de mayo de 1980, al formarse una prominencia en la vertiente septentrional de la montaña (*línea de trazos superior*), y primero de julio de 1980 (*línea de trazos inferior*). La zona en color de la derecha, situada bajo los cortes topográficos, señala la región donde se originaron los miles de seísmos del enjambre producido entre el 20 de marzo y el 18 de mayo. Cuanto más vivo es el color, tanto mayor es la densidad de los terremotos localizados. Los datos pertenecen al Servicio Geológico de los Estados Unidos y al Departamento de Geofísica de la Universidad de Washington.

El coste humano de la erupción puede reputarse grande. Murieron o desaparecieron sesenta y dos personas. Las pérdidas económicas, sobre todo en la industria maderera, rebasaron los mil millones de dólares. De todos modos, el mayor daño ha sido, quizá, de tipo psicológico: la gente que vive al noroeste de la cordillera de las Cascadas veía, en esos volcanes, unos guardianes estáticos y silenciosos, que de pronto se han convertido en amenaza preocupante.

A la enorme avalancha y a la erupción del 18 de mayo les siguieron explosiones menores de ceniza el 25 de mayo, el 12 de junio, el 22 de julio, el 7 de agosto y entre el 16 y 18 de octubre. La extrusión de domos de lava viscosa en el interior del cráter se produjo entre el 13 y el 20 de junio, el 8 y el 9 de agosto y el 18 y 19 de octubre. El volumen total de magma arrojado en las erupciones comprendidas entre el 15 de mayo y el 19 de octubre fue de unos 0,05 kilómetros cúbicos, estimándose su energía en $1,7 \times 10^{17}$ joule, cantidad que es de un orden de magnitud menor que la correspondiente a la erupción del 18 de mayo. Aunque la erupción del 25 de mayo se produjo de noche y con mal tiempo, parece ser que marcó su comienzo un aumento brusco de los temblores volcánicos a las 2 horas 28 minutos de la madrugada. A las 2 horas 45 minutos, el radar del Servicio Meteorológico Nacional indicó que la columna de ceniza había alcanzado 14 kilómetros de altitud. Al cabo de una hora había perdido altura, pero el volcán continuó arrojando nubes de ceniza menores a lo largo de todo el día. Se emitieron nuevas coladas de piroclastos, formados por cenizas y bloques de pumita, que cubrieron parte de la zona del flanco septentrional del Santa Elena que ya habían barrido las primeras coladas de piroclastos. Si bien la erupción de ceniza fue mucho menos voluminosa que la del 18 de mayo, variaron mucho más las direcciones del viento, cayendo una fina capa de ceniza sobre gran parte del occidente de los estados de Washington y Oregon, incluida el área metropolitana de Portland.

Las explosiones del 12 de junio fueron parecidas a la del 25 de mayo. El temblor volcánico empezó por la tarde, produciéndose a las 19 horas 5 minutos la primera emisión de ceniza, que alcanzó una altura de cuatro kilómetros. A las 9 horas 9 minutos de la noche se desencadenó una erupción de ceniza mucho mayor, que alcanzó 15 kilómetros; terminó rápidamente, caída la medianoche. Al día siguiente, la observación desde un helicóptero reveló que otra serie de coladas de piroclastos, de espesores comprendidos entre 2 y 10 metros y temperaturas máximas de 600 grados C, había descendido hacia el Spirit. Tras la erupción del 12 de junio, empezó a formarse, en el cráter de explosión, un domo de lava que el 15 del mismo mes tenía 200 metros de anchura por 40 de alto y estaba surcado por grandes y resplandecientes grietas. El domo continuó elevándose, a razón de seis metros por día, y el 20 de junio había alcanzado su altura máxima de 65 metros.

Estuvo en calma hasta el 22 de julio. Aquella mañana, se registraron pequeños terremotos superficiales, localizados bajo la zona del cráter. El número de terremotos aumentó a lo largo del día pero no se detectó ningún temblor volcánico. De repente, a las 5 horas 14 de la tarde, y con el cielo despejado, propio de un día de verano, empezó a salir una gran nube de ceniza procedente de la montaña. El radar señaló que la parte superior de esta nube alcanzaba los 14 kilómetros. Una segunda nube de ceniza entró en erupción a las 6 horas 25 minutos de la tarde y llegó a una altura de 18 kilómetros en sólo siete minutos 23 segundos, con una velocidad media de 2,2 kilómetros por minuto. El tercer surtidor de cenizas, y el más largo, empezó a las 19 horas 1 minuto y duró más de dos horas, elevándose a una altura máxima de 14 kilómetros.

Los geólogos y los bomberos del Servicio Forestal, avisados del aumento de los terremotos superficiales de pequeña intensidad, habían abandonado la zona, pero las erupciones se siguieron observando desde helicópteros y aviones. Las coladas piroclásticas emitidas por la abertura volcánica durante la segunda y tercera explosiones de ceniza descendieron por la vertiente septentrional del volcán hacia el lago Spirit. Richard Hoblitt, del Servicio Geológico de los Estados Unidos, fue testigo presencial del fenómeno y lo describió de esta manera:

"Estábamos volando de este a oeste, y aproximadamente a una milla al norte de la abertura del volcán, cuando empezó la segunda erupción. Instantes más tarde, durante los cuales el ritmo de emisión del gas aumentó, se formó un surtidor de cenizas hasta alcanzar unos 500 metros por encima de la abertura. A medida que las proyecciones del surtidor se arqueaban y caían en las cercanías de la abertura, producían una colada de piroclastos que descendía rápidamente hacia el norte, alejándose del anfiteatro. Salimos hacia el oeste lo más rápidamente que pudimos." Poca gente que haya visto tan de cerca una colada piroclástica ha sobrevivido para contarlo.

Estas nuevas coladas de ceniza y bloques de pumita tenían un espesor comprendido entre uno y dos metros. Su temperatura se midió al día siguiente, siendo el valor máximo obtenido de 705 grados C, a una profundidad de 1,5 metros. Las nubes de ceniza derivaron hacia el nordeste el 22 de julio, observándose caídas de ceniza de poca importancia en la parte central y occidental del estado de Washington.

La erupción del 22 de julio vino precedida por un cambio en el tipo de emisión gaseosa. Las emisiones de gas cambiaron de nuevo a principios de agosto y los temblores volcánicos empezaron el 7 de agosto, inmediatamente después del mediodía. Advertidos por estas dos señales, los investigadores que se encontraban en las zonas de peligro fueron evacuados. A las 4 horas 26 minutos de la tarde, se iniciaba una erupción que originó una nube de ceniza que alcanzaría los 13 kilómetros de altura. Pequeñas coladas de piroclastos barrieron la parte situada por debajo de la zona de ruptura, en la vertiente septentrional de la montaña, mientras que otras erupciones, más pequeñas aún, continuaron produciéndose al final de la tarde y al anochecer, seguidas de una gran explosión a las 10 horas 30 minutos de la noche. Entre el 8 y el 9 de agosto se formó en el cráter un nuevo domo de lava.

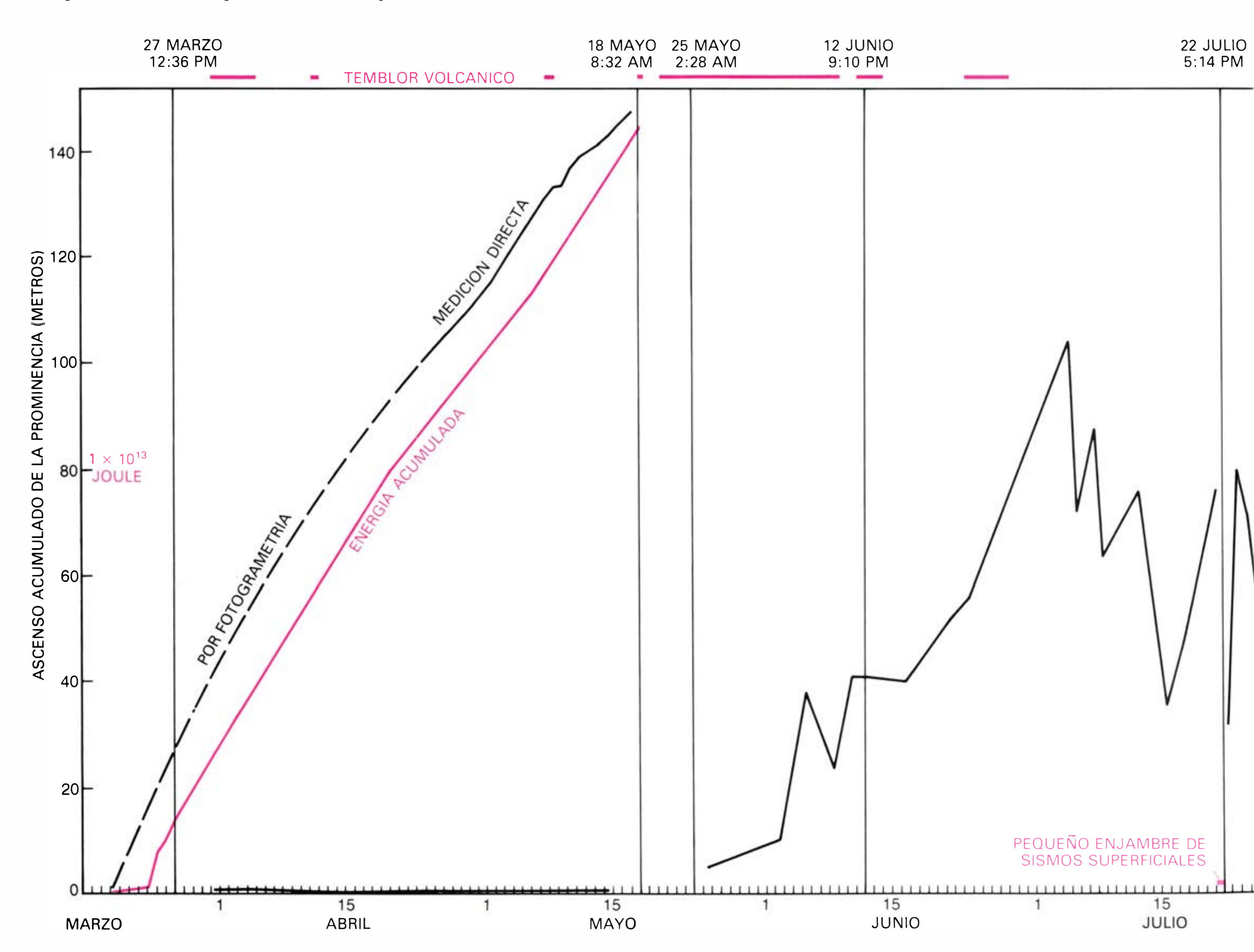
Transcurrieron dos meses en que predominó la tranquilidad. Pero el 16 de octubre volvieron los terremotos de pequeña intensidad, parecidos a los que habían precedido a la erupción del 22 de julio. Cuando el enjambre aumentó, se dio orden de alerta a primeras horas de la noche. A las 21 horas 58 minutos, se produjo una erupción. Cuatro explosiones de ceniza, ocurridas durante los dos días siguientes, lanzaron nubes hasta 14 kilómetros de altura y destruyeron el domo de lava formado en agosto. Algunas erupciones se acompañaron de pequeñas coladas de piroclastos, que descendieron por la vertiente norte de la montaña. Cenizas de colores claros cayeron hacia el sur y sudeste, recordándole a la ciudad de Portland que el monte Santa Elena seguía activo. Entre el 18 y el 19 de octu-

bre apareció un nuevo domo de lava, el mayor de los surgidos hasta entonces.

Las erupciones de los volcanes presentan un carácter explosivo extraordinariamente variable. En las islas Hawai, las efusiones de lava incandescente son arrojadas espectacularmente al aire sin producir daño alguno, y las coladas de roca fundida descienden lentamente por las vertientes desde los cráteres. Completamente distintas son las gigantescas explosiones que destruyen montañas enteras, y que se producen cuando una parte importante de la energía calorífica del magma se convierte en trabajo mecánico. Una de las medidas del carácter explosivo de los volcanes es la naturaleza de sus productos. Las erupciones de tipo efusivo se caracterizan por sus coladas de lava, mientras que las erupciones explosivas se distinguen por lanzar fragmentos como ceniza volcánica, pequeños trozos de roca y bloques. En las islas Hawai, cerca del 98 por ciento de los productos de erupción son lavas efusivas, mientras que en los volcanes que bordean el Pacífico ocurre casi lo contrario, es decir, el 90 por ciento de los productos de las erupciones son fragmentos de roca.

Los volcanes se localizan en tres ámbitos de tectónica peculiar. Los volcanes de cada una de estas zonas difieren notablemente de los de las otras en lo que a su comportamiento explosivo se refiere. Tal como hemos visto, los volcanes del círculo de fuego se levantan a lo largo de las zonas de subducción, donde se produce la convergencia de las placas tectónicas. Estos volcanes son, generalmente, de tipo explosivo. Los volcanes de fisura se localizan en zonas de divergencia de placas y sus erupciones son de tipo más efusivo, especialmente las submarinas profundas. Sin embargo, si la erupción de los volcanes fisurales se produce con poca profundidad de agua, o en la corteza continental, puede entonces ser explosiva. Los volcanes de los focos térmicos, que atraviesan las placas tectónicas, son generalmente efusivos cuando se localizan en zonas con corteza oceánica (Hawai) y explosivos en zonas de corteza continental (Yellowstone).

Los factores clave que determinan el carácter explosivo de los volcanes parecen ser la viscosidad del magma, la cantidad de gas disuelto en el mismo, la cantidad de agua subterránea existente cerca de su boca de salida y la presión superficial. Las explosiones volcánicas se deben a la rápida expansión de gas dentro del magma o en íntimo contacto con él. No se produce una liberación de energía cual si fuera una bomba, sino que, preferentemente, recuerda la explosión de una caldera de vapor. Podemos comparar la viscosidad con la resistencia de la caldera; cuanto mayor sea la viscosidad, tanto mayor podrá ser la explosión potencial. La disponibilidad de gas dentro del magma, o en contacto con él, guarda relación con el volumen de la caldera. La diferencia entre la



CUATRO TIPOS DE ACTIVIDAD asociados con las erupciones, representados en estos gráficos: enjambres de terremotos, formación de la prominencia, emisión del gas anhídrido sulfuroso y temblores volcánicos. Las principales erupciones se indican por líneas verticales. La deformación máxima de la

presión del gas en el magma y la presión circundante se puede comparar con la diferencia de presión a ambos lados de las paredes de la caldera. El vapor de agua es uno de los gases magmáticos principales, pero el anhídrido carbónico desempeña también un papel destacado en la explosión volcánica.

Vamos a considerar dos ejemplos. En el primero tenemos un magma basáltico, con una viscosidad relativamente pequeña y un contenido bajo en gases disueltos. Si el basalto asciende por una fractura hacia el exterior, el anhídrido carbónico y el agua empiezan a separarse de la disolución a medida que disminuye la presión. Las burbujas aumentan de volumen, venciendo las débiles fuerzas de la viscosidad, y producen la expansión del magma espumante. Este proceso adquiere su máxima importancia en los metros finales, antes de que el magma alcance la superficie, dando origen a un surtidor de lava incandescente en efervescencia, en la abertura de salida. Si el mismo basalto surgiera bajo dos kilómetros de agua de mar, no bastaría la caída de presión para que el agua disuelta en el magma se liberara de la solución. Seguramente, se desprendería algo de anhídrido carbónico, pero la expansión de la mezcla de este gas con el magma sería pequeña y se produciría una salida tranquila por su abertura submarina. Si la profundidad de agua fuera pequeña, el carácter explosivo podría ser variable, en función del ritmo con que se produjera la mezcla de la lava efusiva a elevada temperatura con el agua circundante.

El segundo ejemplo es un magma silíceo con una viscosidad relativamente elevada y alto contenido en gases disueltos. Si este magma dacítico asciende hacia la superficie, el gas que se desprende de la solución forma burbujas. Sin embargo, debido a la elevada viscosidad de la fusión, las burbujas no aumentan de volumen hasta sus dimensiones de equilibrio, porque tienen una sobrepresión interna estimada en varios centenares de atmósferas. Cuando se forman suficientes burbujas y la presión externa es lo bastante baja, se rompe el magma en fragmentos que salen aceleradamente por la abertura gracias a los gases liberados de forma explosiva. El enfriamiento de los gases en expansión solidifica los fragmentos de magma que son emitidos en una emulsión en forma de partículas a elevada temperatura, pero sólidas. Estas mezclas fluidificadas constituyen nubes de erupción o coladas piroclásticas, según sea su densidad y la velocidad y dirección de su eyección.

prominencia y la energía sísmica están relacionadas con el período anterior a la erupción del 18 de mayo. (Datos del Servicio Geológico y el Departamento de Geofísica de la Universidad de Washington.)

Uno de los casos más afines a la erupción del monte Santa Elena fue la del Bezymianny, en la península de Kamchatka. Ocurrió entre 1955 y 1956. El vulcanólogo ruso G. M. Gorshkov ha distinguido cinco fases en esta erupción: 1) del 29 de septiembre al 21 de octubre de 1955, enjambre de sismos anteriores a la erupción; 2) desde el 22 de octubre hasta el 30 de noviembre de 1955, importantes erupciones de ceniza; 3) desde el 1 de diciembre de 1955 hasta el 29 de marzo de 1956, erupciones moderadas de ceniza, con levantamiento generalizado del domo de lava antiguo; 4) marzo de 1956, gigantesca explosión con salida lateral; 5) abril de 1956, formación de un nuevo domo de lava en el cráter de explosión.

El volumen de roca movilizado y la zona afectada por la erupción del Bezymianny, el 30 de marzo de 1956, son similares a los de la erupción del Santa Elena el 18 de mayo de 1980. Las diferencias más significativas residen en que, en la erupción del Bezymianny, el enjambre sísmico que precedió a la primera erupción de ceniza duró 23 días, mientras que en el monte Santa Elena sólo fue de 7 días, y en que las erupciones preliminares duraron 160 días, frente a 53, habiéndose producido explosiones de ceniza mayores, en la primera fase de las explosiones preliminares ocurridas en el Bezymianny. En este último lugar, la explosión principal emitió sobre todo rocas constituidas por magma nuevo, contrariamente a lo ocurrido en la avalancha y explosión del 18 de mayo en el monte Santa Elena, donde menos de la quinta parte de las rocas lanzadas al exterior era de magma reciente. Quizá sea ésta la causa principal de que el Bezymianny haya permanecido en actividad desde 1956. Su última erupción importante ocurrió en 1979.

El monte Santa Elena ha permitido ensayar adecuadamente las técnicas de

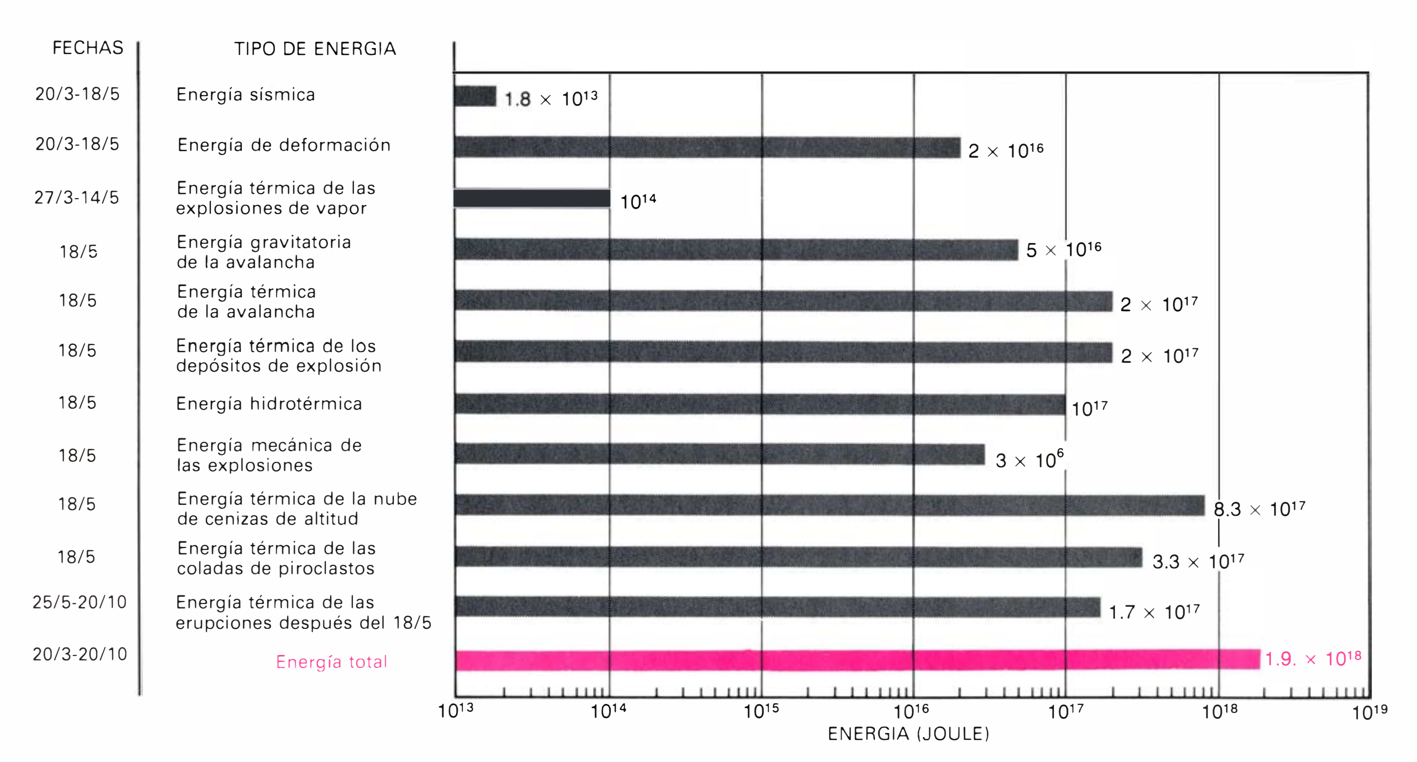
predicción de erupciones volcánicas. Para ello no hay que ceñirse a un solo aspecto, sino que deben valorarse todos los factores posibles y, luego, interpretarlos a la luz de la experiencia geológica. Recordemos, entre estos factores, las estadísticas de las erupciones ocurridas a lo largo de la historia y la reconstrucción de las estadísticas de las erupciones prehistóricas, mediante datación y cartografía geológicas. De las técnicas geofísicas citemos la vigilancia (en los volcanes activos o potencialmente activos, o en sus cercanías) de la sismicidad, la deformación de la superficie topográfica, los campos magnético y eléctrico y las temperaturas. Las técnicas geoquímicas incluyen la vigilancia del volumen y composición de los gases, líquidos y sólidos emitidos por los volcanes. Las repetidas observaciones visuales realizadas desde el suelo y desde el aire proporcionan importantes datos sobre la evolución de la actividad volcánica. La mayoría de los volcanes del mundo ni siquiera reciben este tipo de atención básica; por descontado, tampoco son vigilados por los instrumentos de gran perfección disponibles hoy en día.

En el monte Santa Elena, unos métodos funcionaron mejor que otros. Los datos de su actividad histórica nos sitúan las erupciones durante el período comprendido entre 1831 y 1857, lo cual, si bien permite clasificar al volcán como activo, nos informa sobre un lapso de tiempo demasiado corto para que pueda tener significado estadístico. La cartografía geológica y las dataciones permitieron establecer 20 períodos eruptivos, con emisión de productos diversos, durante los últimos 4500 años, demostrando la existencia de una serie de caracteres distintivos importantes. Los intervalos de inactividad determinaron dos poblaciones: la de 100-150 años y la de 400-500 años. Los dos últimos intervalos de inactividad, anteriores a 1800, fueron del tipo de corta duración. Los materiales eruptivos demostraron también que eran frecuentes las explosiones de ceniza y piroclastos y que las erupciones históricas del monte Santa Elena afectaban a grandes zonas. Fue este tipo de análisis el que llevó a Crandell y a Mullineaux a predecir la peligrosidad del volcán.

En 1980, la vigilancia sísmica del enjambre de terremotos en aumento localizados bajo el monte Santa Elena alertó, con una semana de anticipación, sobre las pequeñas erupciones explosivas. La fotogrametría y las medidas electro-ópticas de distancia dieron cuenta de la impresionante deformación de la superficie topográfica asociada a la prominencia creciente de la vertiente norte de la montaña. No hubo cambios en el ritmo de la sismicidad o en el de la deformación del suelo antes de la erupción del 18 de mayo, pero el hecho de que estos fenómenos prosiguieran fue para todos un aviso de que algo importante estaba ocurriendo bajo el volcán. Este aviso permitió al gobernador Ray y al Servicio Forestal de los Estados Unidos mantener su orden de clausura de la zona, a pesar de las continuas peticiones de libre acceso a ella. Su firmeza salvó millares de vidas.

En los meses siguientes al 18 de mayo de 1980, se previeron con éxito las pequeñas erupciones explosivas, pero no por ello menos significativas, del 12 de junio, 22 de julio, 7 de agosto y 16 de octubre. Varias horas antes de las erupciones del 12 de junio y del 7 de agosto se produjeron temblores volcánicos, y también horas antes de las erupciones del 22 de julio y del 16 de octubre se registraron pequeños terremotos anormales debajo mismo de la montaña. Una serie de anomalías en las características de las emisiones de gases y de las deformaciones menores de la superficie topográfica precedieron, igualmente, en horas o en días, a algunas de las erupciones producidas con posterioridad al 18 de mayo.

Cualquier cambio de las características globales observado por los aparatos de vigilancia despierta sospecha y la experiencia ayuda a valorarlo. Se han registrado falsas alarmas, inevitables por lo demás en los temas de predicción aleatoria. La tecnología no ha resuelto aún el viejo problema de discernir el grado de certeza que nos faculta para dar, sin precipitación, la señal de alarma real.



SEIS TIPOS DE ENERGIA liberada en las erupciones. Las fechas de la liberación de energía se dan en la columna de la izquierda. La energía sísmica, la energía de deformación y la energía térmica de las explosiones de vapor de agua fueron liberadas durante el período anterior a la erupción del 18 de mayo. Se ha seguido una escala logarítmica para las barras horizontales. La última barra (*en color*) refleja la energía total liberada hasta octubre.

Ciencia y sociedad

La investigación biológica en la Edad Media

En el tránsito entre las edades Antigua y Medieval, Agustín de Hipona establece en su escrito *De doctrina christiana* lo que, siguiendo la terminología introducida en historia de la ciencia por I. Lakatos, podemos denominar un "programa de investigación". Su objetivo último no era propiamente científico, sino teológico, el logro de la *sapientia divina,* y el método consistía en el análisis minucioso del "libro de Dios", es decir, de la Escritura. Ahora bien, Agustín era consciente de que resultaba imposible el logro de la *sapientia* sin el concurso de la *scientia,* que no era fácil la intelección cabal de lo que el libro de Dios dice sin el auxilio del "libro de la creación", al que la Escritura hace continuas referencias. Para la conquista de la *sapientia divina* es muy útil el concurso de la *scientia naturalis,* que tan bien habían compilado los enciclopedistas romanos, Varrón, Cicerón, Lucrecio, Plinio, Celso, etcétera. He aquí el programa científico agustiniano: reconstruir la obra de Varrón para uso de cristianos, escribir una enciclopedia que, resumiendo todos los conocimientos existentes sobre las cosas creadas, permitiera entender adecuadamente la palabra de Dios.

Este programa de investigación intentaron desarrollarlo bastantes autores, como Casiodoro (s. VI), Isidoro de Sevilla (s. VII), Beda el Venerable (s. VIII), Rhabano Mauro (s. IX), Alejandro Neckam (s. XII) y, ya en el XIII, es decir, cuando el "núcleo firme" (*hard core*) del programa agustiniano estaba siendo sometido a intensa crítica, hicieron un último esfuerzo por consolidarlo un cierto número de autores que, bien mediante el recurso a una erudición enorme, cual fue el caso de Vicente de Beauvais, bien mediante el uso pragmático de los libros naturales de Aristóteles, como sucedió en Tomás de Cantimpré, intentaron establecer un "círculo protector" de ese núcleo ya herido de muerte. De ahí que las obras de estos últimos sean, a la vez que el canto de cisne del género, su certificado de defunción. En el libro XIX de su enciclopedia *De naturis rerum,* editada por L. García Ballester, expone así Tomás de Cantimpré su programa de investigación: "Han pasado ya catorce o quince años desde que acometí con toda diligencia la realización del libro *De naturis rerum,* siguiendo el consejo de San Agustín en su obra *De doctrina christiana,* que continuamente tengo ante mis ojos, donde afirma la gran utilidad que se derivaría si alguien emprendiera el trabajo de reunir en una sola obra lo dicho sobre la naturaleza de las cosas y especialmente de los animales".

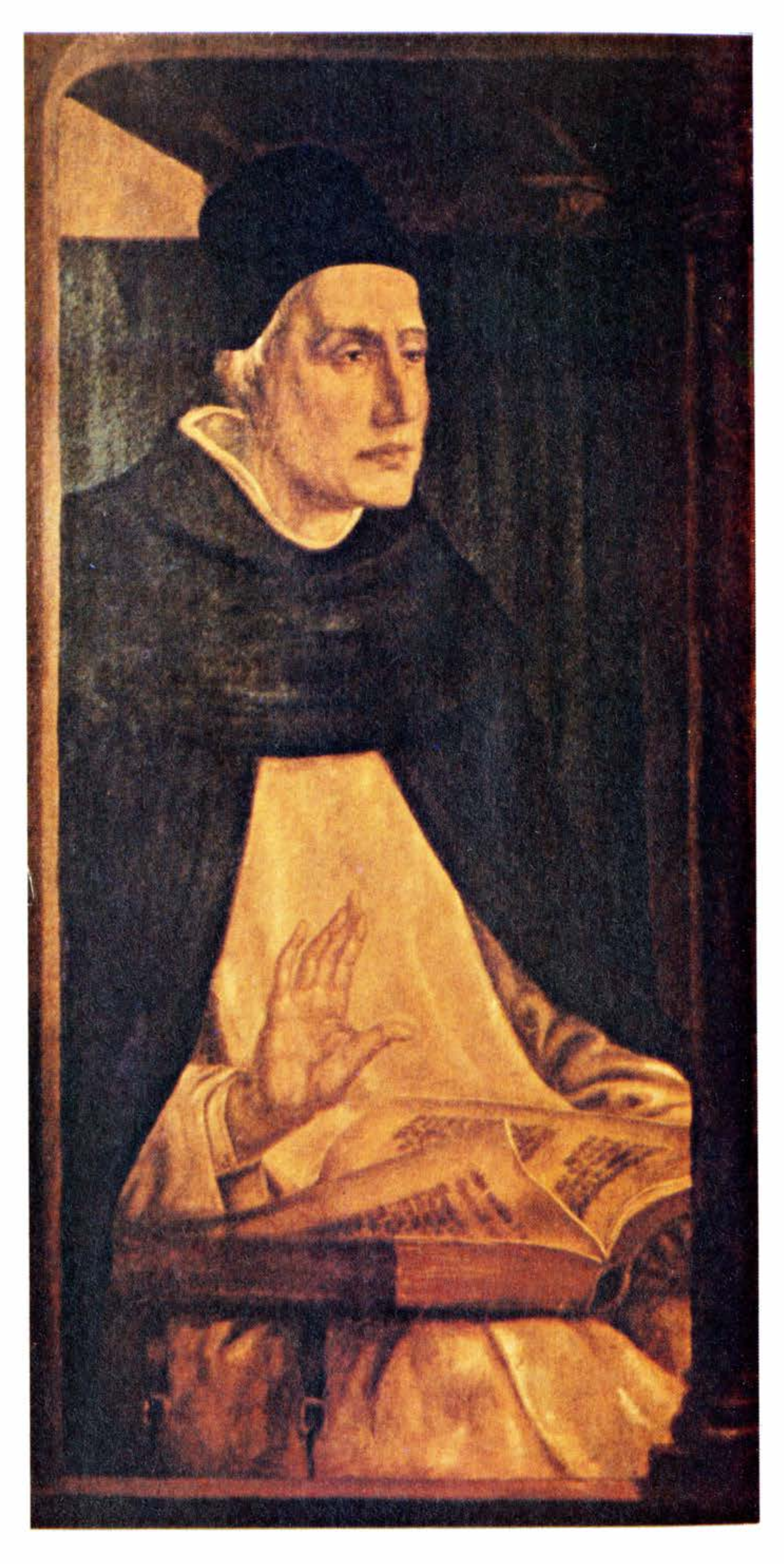
El orden expositivo de este género de escritos suele guardar siempre un cierto parecido con el utilizado por Varrón: historia, geografía, botánica, zoología, mineralogía, astronomía, etcétera. Naturalmente, no era fácil cubrir con suficiencia un programa tan amplio, y muchos habían de renunciar a tal o cual grupo de disciplinas en favor de otro o de otras. Dado nuestro propósito, aquí nos interesan los desarrollos relacionados con la biología. Un ejemplo del método expositivo seguido en estos libros lo ofrecen las siguientes líneas del ya citado libro de Tomás de Cantimpré: "En primer lugar he de tratar sobre la anatomía del cuerpo humano, sus afecciones y los remedios a las mismas. Sigue un breve y provechoso tratado sobre el alma cuyo poder dilucidó con total claridad Agustín, doctor incomparable, en su libro *De anima et spiritu.* Después, los tipos de hombres monstruosos y las razones de su origen. Luego, la naturaleza de las bestias, las aves y los monstruos marinos, y también los peces, serpientes, gusanos, árboles, hierbas, ríos, piedras, metales y humores del aire; y además, sobre los siete planetas, los fenómenos del aire, sobre los movimientos del sol y la luna, y los eclipses de los mismos; y, por último, sobre los cuatro elementos".

Los "bestiarios" o tratados de animales escritos con esta mentalidad tienen un aire muy singular que les viene conferido por su intención, siempre moralizante y simbólica. En la Antigüedad se había explotado ya con gran éxito este género, como lo demuestran las fábulas de Esopo, las *Quaestiones naturales* de Séneca y el anónimo alejandrino del siglo II conocido con el nombre de *Physiologus.* Cada animal es el símbolo de una virtud o un vicio, y de su vida siempre es posible extraer alguna conclusión moralizante o de edificación. Así, el ave fénix es el símbolo de Cristo crucificado, y la hormiga-león, nacida por concurso del león y de la hormiga, es un ser contradictorio y miserable, como el hombre que intenta contentar a la vez a Dios y al diablo (A. C. Crombie). Del mismo modo, Tomás de Cantimpré dice al comienzo de su enciclopedia que intenta reunir las noticias dispersas en los autores clásicos "sobre la naturaleza de las cosas creadas y sus propiedades, dignas de memoria y apropiadas a su significación moral". Señalando ya desde el prólogo de la obra el carácter moralizante de su empeño, Tomás de Cantimpré no hace otra cosa que explicitar uno de los postulados básicos o nucleares del programa agustiniano. Lo original en Cantimpré es otra cosa, su intento de proteger ese núcleo con los propios argumentos que en esos años se estaban esgrimiendo en su contra, es decir, con las obras biológicas de Aristóteles. Y escribe: "Dice Aristóteles en el libro XI de su obra *De animalibus* que, ya se trate de los más nobles como son los celestiales, ya de los innobles como son los animales, su conocimiento –los que sean capaces de él– procede del gran amor que les profesa (la divinidad). Así pues, por esta razón debemos prestar atención a la diversidad de criaturas y deleitarnos en el artífice que las creó puesto que la ciencia del que actúa se manifiesta en su obra. Debido a ello sostenemos, y no nos resulte gravoso, que en la naturaleza de los animales viles, y en la de todas las cosas naturales, hay algo admirable. Las cosas naturales son nobles ya que ningún ser ha sido creado sin sentido ni por azar sino para complementar y, por tanto, desempeña alguna función digna de consideración y tiene un puesto noble".

Hacia mediados del siglo XII se produce un acontecimiento que en muy pocas décadas iba a dar al traste con ese programa agustiniano de investigación biológica. Hasta el primer tercio de esa centuria no fueron conocidos en Occidente más tratados aristotélicos que algunos capítulos del *Organon* que Boecio tradujo y recopiló en la denominada *Logica vetus.* Sólo en ese siglo comenzaron a conocerse los demás libros de la lógica aristotélica, traducidos y compilados bajo el nombre genérico de *Logica nova.* Fue el principio de la recepción del "nuevo Aristóteles", cuyo plan, siguiendo el del *Kitab al-Sifa* de Avicena, se encuentra ya claramente formulado en la primera fase de la escuela de traductores de Toledo, lo que H. Schipperges ha denominado *Frühtoledo* (aproximadamente 1120-1140). Tras la *Logica* o *Collectio prima* se sitúan los *Naturalia* o *Collectio secunda,* seguida ésta de la *Collectio tertia* o ma-

temática y, finalmente, de la *Collectio quarta,* dedicada a la *philosophia prima* o metafísica. Este es el sistema aristotélico de la ciencia que, reelaborado por los filósofos árabes del Oriente, fue recuperado por Toledo, donde desde fines del siglo XI hasta principios del XIII se conocieron una buena parte de las obras de Aristóteles, y donde además se desarrolló una gran actividad tendente a fijar de forma definitiva el orden aristotélico de las ciencias. (A tal fin se traducen el *Catálogo de las ciencias* de Alfarabi, el libro de Avicena titulado *División de las ciencias,* muy influido por el anterior, y el de Algazel, *Intenciones de los filósofos,* directamente redactado sobre la base del de Avicena. Además, el primero fue refundido y extractado por Domingo Gundisalvo en sus obras *De divisione philosophiae* y *De scientiis,* y traducido más tarde de modo completo bajo el título de *De scientiis* por Gerardo de Cremona.) Desde Toledo, este catálogo aristotélico de las ciencias se extendió por toda Europa y reorientó las discusiones científicas de los siglos XIII y siguientes en una dirección nueva. No se trataba de un mero añadido al saber tradicional sino de algo que introdujo una novedad radical en las discusiones científicas, un nuevo "programa de investigación", distinto del agustiniano.

El lugar asignado a las ciencias biológicas es la *Collectio secunda naturalium.* Unas líneas de Avicena en el prólogo al libro sexto de esa colección da idea exacta de su contenido: "En el primer libro expusimos ya las cosas que son comunes a los seres de la naturaleza. A éste le añadimos un segundo libro, dedicado al conocimiento de los cuerpos y de las formas, así como de los movimientos primeros en tanto que naturales, y comprobamos las disposiciones de los cuerpos, de aquellos que se corrompen y de los que no. Tras esto ordenamos el tratado sobre la generación y la corrupción y de sus sujetos. Después añadimos las cuestiones sobre las acciones de las cualidades primeras, y de las pasiones y complexiones que se generan en ellas. Era preciso que tras esto hablásemos de las cosas generadas; y como las cosas inanimadas e insensibles y que carecen de movimiento voluntario son las primeras y más aptas para ser generadas desde los elementos, hemos hablado de ellas en el libro quinto. Tras lo cual faltaba a la ciencia natural el considerar las cosas vegetales y animales. Los cuerpos vegetales y sensibles son aquellos cuya esencia se constituye de la forma, que es el alma, y de la materia, que es el cuerpo y sus



Alberto Magno critica el estudio de los animales con fines moralizantes y teológicos e introduce en la observación de la naturaleza un espíritu nuevo

miembros; y como el mejor conocimiento posible de ellos es aquel que se adquiere por su forma, hemos elegido el hablar primero del alma... y éste es el libro sexto, al que añadiremos el séptimo, que es el tratado de las disposiciones de los vegetales, al que seguirá el libro octavo, sobre las disposiciones de los animales, y donde culminará la ciencia natural".

Si se compara este programa científico con el agustiniano, vemos su gran diferencia. Este no se halla basado en las obras de los enciclopedistas romanos sino en las obras naturales de Aristóteles, ordenadas generalmente así: *Physica, De caelo et mundo, De generatione et corruptione, De meteorologicis, De animalibus* (que comprende *Historia animalium, De partibus animalium* y *De generatione animalium*), *De anima* y *Parva naturalia.* Pero, además, el método de análisis es distinto. La diferencia tajante que se establece en el siglo XIII entre causas primeras y segundas aleja del horizonte de los *naturalia* toda consideración de tipo teológico o religioso. De lo que ahora se trata es, como ya hizo el propio Aristóteles, del conocimiento de los seres naturales por lo que ellos son, sin ningún tipo de justificación extrínseca. Esto se advierte ya con claridad en un escritor muy influenciado por el espíritu toledano y por el ambiente cultural de la corte de Alfonso X el Sabio, Juan Gil de Zamora, autor de una *Historia naturalis* compuesta según el nuevo espíritu. Pero donde el nuevo programa de investigación alcanza su cima es en Alberto Magno, figura tan gigantesca que Gilson no ha dudado en caracterizar la novedad y la amplitud de su empresa intelectual como "ideal pantagruélico del saber". Alberto Magno, que se sabe innovador y hace gala de ello en sus escritos, critica el estudio de los animales con fines moralizantes y teológicos e introduce en la observación de la naturaleza un espíritu realmente nuevo. "No se pueden hacer silogismos sobre naturalezas concretas, de las que sólo la experiencia (*experimentum*) da certeza", dice en *De vegetabilibus et plantis.* La biología, concretamente, debe pasar de las manos de los predicadores y de los místicos a la de los naturalistas. Con este talante compone su tratado *De animalibus,* cuyos diecinueve primeros libros exponen los tres grandes tratados aristotélicos que los medievales conocieron con el nombre genérico de *De animalibus,* dedicando los seis últimos a la discusión original de varios temas biológicos y a la descripción de animales concretos. Frente a las tesis biológicas de corte galénico, que habían venido siendo las tradicionales, como la contribución de ambos progenitores a la forma del nuevo ser o como el cerebrocentrismo, Alberto Magno defiende resueltamente las opiniones aristotélicas, la procedencia masculina de la forma del embrión, el cardiocentrismo, etcétera. Actuando así, Alberto Magno lleva hasta sus últimas consecuencias el nuevo espíritu toledano. Recuérdese, por ejemplo, la importancia que en el desarrollo del cardiocentrismo tuvieron los tratados *De motu cordis,* el primero de los cuales lo compone precisamente un hombre relacionado con el programa toledano, Alfredo de Sareshel. O la cantidad de comentarios que a partir de entonces se escriben al *De generatione animalium* de Aristóteles, a fin de rechazar la teoría galénica.

Pero la vida de este segundo programa de investigación iba a ser muy efímera. Nunca resultó posible establecer con claridad sus postulados básicos o nucleares, única forma de hacerlo estable. De este modo, la Edad Media volvía a revivir la ambigüedad que se halla en la base del propio sistema aristotélico, y que ha servido a Jaeger para establecer la línea evolutiva de su pensamiento, desde el conceptismo platonizante de la primera época al naturalismo empirista de la última. Los libros biológicos atribuidos a Aristóteles debieron escribirse en el último período, y suponen en alguna medida –sólo en alguna medida– el abandono de la vía de la predicación, propia de los escritos anteriores, por la vía de la naturaleza, de la generación y la corrupción. De ser esto cierto, su autor habría evolucionado desde una filosofía preponderantemente "lógica", basada en la idea de la sustancia segunda o universal, hacia una filosofía "física", que fija su atención en la sustancia primera o individual, esto es, en la cosa real concreta (*De partibus animalium* I,5). Ciertamente, en las obras de Aristóteles es imposible acceder al término de este proceso, razón por la cual la filosofía que contienen es siempre en última instancia "lógica" o "predicativa". Es, de hecho, el modo como el aristotelismo ha tenido auténtica vigencia histórica. El avicenismo y el albertinismo participan también de esa mentalidad, y por tanto establecen el problema lógico de la predicación como uno de los postulados básicos del "núcleo firme" del programa. Ahora bien, el hecho de que en el propio *Corpus aristotelicum* fuera muy difícil defenderlo como tal demuestra su fragilidad. En el caso que estamos estudiando esto se confirma con toda evidencia. La autonomía que en el avicenismo y el albertinismo adquieren los saberes naturales, en especial los biológicos, reforzó de tal manera la vía de la *phýsis* que pronto acabó volviéndose contra el propio núcleo del sistema, negando el carácter central de la predicación lógica. Dicho en términos más simples, el nuevo papel concedido a la "experiencia" acabó cuestionando la primacía del *lógos*. Si éste hubiera sido sustituido por aquélla, se habría establecido un nuevo "núcleo firme" y, con ello, un programa distinto de investigación. En la Edad Media no se llegó a tanto. Pero sí se hizo una crítica implacable del programa avicenista y albertiniano, obra casi simultánea de dos movimientos, el averroísmo y el nominalismo.

El primer movimiento que contribuyó de modo decisivo a desarticular el programa avicenista y albertiniano fue el averroísmo latino. Su efecto puede calificarse de paradójico, ya que el averroísmo intentó ser la restitución del aristotelismo a su primitiva pureza. Pero el hecho de que esta ideología se extendiera principalmente en las comunidades judías, sumado al de que en éstas hubiera un gran número de personas dedicadas a los estudios biológicos y médicos, hizo que tales estudios alcanzaran en ellas una importancia excepcional. Añádase otro dato de no menor trascendencia, a saber, el mayor despego de las comunidades judías respecto de la disciplina canónica y, por tanto, la existencia de una mayor libertad ante las repetidas prohibiciones papales relacionadas con los estudios biológicos, lo que les permitió llevar a cabo experi-

mentos inusuales y de singular importancia, como la disección de cadáveres humanos con fines estrictamente científicos.

Rescatando al biólogo de la tutela canónica y concediéndole omnímoda libertad para investigar, los averroístas latinos estaban sentando las bases de una revolución técnica y política que en muy pocos años habría de demostrar los errores biológicos de las obras de los antiguos y la necesidad de conceder un papel nuevo y más importante a la *experientia* y el *experimentum.* Pero junto a ella hay que colocar la otra gran revolución, la propiamente filosófica, llevada a cabo por los nominalistas, y en particular por Guillermo de Ockam. El fue quien dio un giro completo al aristotelismo, al conceder primacía total a la sustancia primera sobre la segunda, a la naturaleza sobre la predicación, a la experiencia sobre la razón. El conocimiento experimental o intuitivo es, dice Ockam, el único capaz de dar evidencia inmediata, en tanto que el conocimiento abstracto o lógico sólo la tiene en tanto que recibida de aquél, siendo su evidencia, por ello, mediata. Ockam no identifica ciencia con experimento; la ciencia es algo más, conocimiento universal lógicamente estructurado. Pero lo que aprueba o reprueba una formulación científica es siempre el hecho experimental, que adquiere así el carácter de prueba. *Perfecta cognitio intuitiva est illa de qua dicendum est quod est cognitio experimentalis, et ista cognitio est causa propositionis universalis quae est principium artis et scientiae.*

El averroísmo y el nominalismo eran movimientos entre sí heterogéneos. De ahí que en vez de ofrecer de modo natural e inmediato el conjunto de postulados nucleares de un nuevo programa de investigación, entraran entre sí en viva polémica. No podemos entrar en su análisis. Pero sí hay que decir que ella contribuyó poderosamente a la constitución de los nuevos programas de investigación biológica en los comienzos del mundo moderno. (Diego Gracia.)

Problemas de la hortofruticultura

Empecemos por la citricultura. Entre los problemas más graves que tiene planteados destacan los siguientes. En primer lugar, los causados por virus patógenos que afectan gravemente a la productividad y a la longevidad de los árboles. Por la importancia de daños causados, la tristeza de los cítricos es, sin duda, la más peligrosa. Esta enfermedad, que se manifestó en España por primera vez en 1958, afecta en la actualidad, con mayor o menor intensidad, al arbolado de unas 80.000 hectáreas. Se estima que, desde su aparición, ha causado la muerte de unos cuatro millones de árboles.

Con objeto de paliar los daños causados por la tristeza se ha recurrido al empleo de patrones tolerantes que ha resultado ser una solución parcial. Durante los últimos años, los trabajos se han orientado hacia la búsqueda de otra forma de lucha, consistente en la inoculación de los árboles con razas de virus causantes de la tristeza de los cítricos poco virulentas, al objeto de que los protejan frente al posterior ataque de razas más virulentas. Esta forma de lucha se está utilizando ya con algún éxito en Brasil. Esto exige un profundo conocimiento de las razas del virus de la tristeza existentes en las plantaciones españolas. Es necesario ahondar también en el mecanismo y la velocidad de propagación de la enfermedad e insectos vectores de la misma, si se quiere luchar con eficacia contra su avance o extensión en el área citrícola.

Además de la tristeza, existen otras enfermedades causadas por virus, viroides y micoplasmas que constituyen graves problemas para la citricultura, puesto que, en general, todas ellas inciden sobre la capacidad productiva del arbolado, reduciéndola, y haciendo más sensibles las plantaciones a otras plagas y enfermedades; entre ellas cabe destacar las siguientes: exocortis, psoriasis, xiloporosis, impietratura y stubborn.

Siguen los problemas derivados de una utilización poco adecuada de los fertilizantes y del agua de riego. La fertilización de los cítricos que se practica en España carece de bases científicas y el rendimiento que se obtiene de la misma es escaso; esta situación persiste año tras año, a pesar de su gran influencia, no sólo en la productividad del arbolado, sino también en la calidad de los frutos. Aunque en la mayoría de los países productores se han realizado gran número de estudios sobre la fertilización, los datos disponibles no bastan para establecer pautas racionales de abonado. Los árboles productores de frutos cítricos pasan, en su ciclo vegetativo anual, por determinados períodos: brotación, floración y cuajado de los frutos; a lo largo de los cuales aumentan sus necesidades de elementos nutritivos, especialmente de nitrógeno y fósforo que, si no son adecuadamente satisfechas, baja de forma notable la producción. Por ello es necesario conocer las épocas óptimas de abonado, al objeto de que la fertilización aportada satisfaga las necesidades totales del árbol y las demandas temporales o estacionales originadas por los períodos críticos citados.

Cuando los fertilizantes se aplican al suelo, una parte de los mismos puede perderse por volatilización, fijación en el suelo o lixiviación por las aguas de riego. Conviene cuantificar estas pérdidas y conocer la fracción del abono aplicado que es útil a las plantas. Dominando esos detalles, será posible realizar una fertilización más racional e incluso conseguir algún ahorro de fertilizantes, lo que habrá de contribuir a mejorar la rentabilidad. Intimamente relacionado con el problema de la fertilización se encuentra el problema del riego; la aplicación del riego tal como se realiza en citricultura (inundación superficial) presenta graves inconvenientes, tales como consumo excesivo de agua, baja eficacia, arrastre de fertilizantes y contaminación de aguas.

Tenemos, en tercer lugar, los problemas debidos a la falta de conocimientos básicos sobre la fisiología de la fructificación y maduración de los frutos. Los cítricos son plantas de floración abundante. La competencia que ejercen entre sí el gran número de frutos, que la planta soporta en el período mayo-julio, motiva una masiva caída de aquéllos después del cuajado, caída que, en algunas ocasiones, es la determinante de la baja productividad del arbolado. Por otra parte, nuestros conocimientos sobre los procesos metabólicos asociados al crecimiento del fruto, así como los mecanismos de acción de los factores hormonales sobre ello, son deficientes y escasos. Si logramos profundizar más en la fisiología de la maduración de los frutos, podremos diseñar con mayor precisión los tratamientos de desverdización a los que se han de someter, prácticamente, todos los frutos que se exportan durante los dos primeros meses de campaña.

Detengámonos brevemente en los problemas causados por insectos-plaga, que han de combatirse con insecticidas tóxicos con las consiguientes secuelas derivadas de esta práctica, tales como la aparición de razas de insectos resistentes a insecticidas y la contaminación de los frutos que comportan los residuos de aquellos tóxicos. Las plantaciones citrícolas están sujetas a la acción de múltiples insectos patógenos que disminuyen la producción y restan calidad al fruto. Para combatirlos se recurre con éxito relativo al empleo de insecticidas. El desarrollo de resistencia a los insecticidas por parte de los insectos y la contaminación de los frutos debida a los restos de los insecticidas que que-

dan sobre ellos constituyen problemas que es necesario evitar. Para ello se impone llevar a cabo estudios que permitan establecer programas de lucha integrada que eviten al máximo el empleo de insecticidas. Sería conveniente también investigar nuevos insecticidas específicos contra las plagas que se desea combatir e inocuos para los mamíferos y las otras poblaciones de insectos.

Para cerrar el capítulo de los problemas que se plantea en los cítricos, consideraremos los provocados por las enfermedades que afectan al arbolado y a los frutos, antes o después de la recolección. Los microorganismos fitopatógenos, hongos principalmente, constituyen un grupo importante de agentes productores de enfermedades de los cítricos que son causa de cuantiosas pérdidas. Estos patógenos afectan en unos casos al arbolado y a los frutos antes de la recolección y, en otros, a los frutos después de la recolección. Entre los hongos que afectan al arbolado destacan los del género *Phytophthora* (causantes de gomosis, podredumbre del cuello y raíces, así como del aguado o podredumbre marrón de los frutos), *Alternaria* (causante de manchas y podredumbre de frutos) y la especie *Colletotrichum gleoesporoides* (causante de la enfermedad denominada antracnosis, que afecta al tejido foliar y a frutos). Entre los hongos que atacan a los frutos después de la recolección destacan los pertenecientes a los géneros *Penicillium* y *Alternaria*. La falta de medios de lucha eficaces contra algunas de las enfermedades fúngicas que afectan a los cítricos requiere un planteamiento de investigación que resulte en formas de lucha aptas para combatirlas.

¿Qué ocurre con los frutales distintos de los cítricos? Presentan, también ellos, una gran diversidad de problemas. En este trabajo vamos a señalar solamente los más importantes, teniendo en cuenta que la mayoría de los apuntados en el caso de los cítricos se repiten en las restantes producciones para la exportación. En los frutales productores de frutos secos importan los problemas de exudaciones gomosas del almendro (causadas por hongos del género *Fusicocum*). En el avellano sobresale la falta de portainjertos aptos para el cultivo de esta especie en las condiciones ecológicas características de la misma y que, al propio tiempo, sean resistentes a enfermedades y de fructificación precoz. El desarrollo de este frutal exige, como paso previo, la búsqueda de un sistema nuevo de plantación que permita abordar el cambio de las actuales plantaciones viejas con garantía de éxito.

PRODUCTO	IMPORTACIONES			EXPORTACIONES		
	1975	1976	1977	1975	1976	1977
Bulbos, cebollas, tubérculos, raíces, brotes y rizomas en reposo vegetativo, en vegetación o en flor	719	662	663	820	99	52
Otras plantas y raíces vivas, incluidos esquejes e injertos	3767	3811	4071	5202	8475	9952
Patatas para siembra	41605	42684	30393	10	131	4142
Legumbres de vaina seca	9	11	6	22	22	—
Trigo	721	1091	1040	14	129	5100
Centeno	503	—	—	11	—	11
Cebada	964	773	417	39	264	2467
Avena	30	5	1	38	30	854
Maíz	1624	1710	450	725	538	775
Sorgo	6405	306	354	77	41	58
Algodón	1057	529	837	—	103	53
Cacahuete	22	29	20	—	—	14
Soja	328	6006	39	—	—	—
Lino	—	12	—	—	—	—
Cáñamo	156	23	23	—	—	—
Flores	62	13	9	2	15	4
Pratenses	1350	1187	803	2329	1622	1599
Hortícolas	87	81	88	48	39	60
Forrajeras	—	18	15	4	35	41
Remolacha azucarera	2961	3567	2374	6	3	—

Balanza de intercambios españoles en semillas selectas, en toneladas métricas
Fuente: Anuario Estadístico de la Producción Agraria 1977.
Ministerio de Agricultura

Dentro del grupo de los frutales de hueso recordemos los problemas de la estacionalidad de la producción del albaricoque y el de las enfermedades causadas por micoplasmas, que acortan fuertemente la vida productiva media del arbolado. En el caso del ciruelo se impone corregir la vecería crónica, que resta potencialidad productiva. El inconveniente mayor asociado a los frutales de pepita quizá sea el de la superproducción. El estudio de nuevas formas de industrialización de manzanas y peras puede contribuir a mejorar la situación de las plantaciones de estos frutales en nuestro país.

El mercado potencial para las frutas tropicales ofrece perspectivas muy favorables, pero el cultivo de estos frutales, aguacate y chirimoyo, principalmente, se enfrentan a múltiples problemas, tales como falta de conocimientos sobre la adaptación de los portainjertos y variedades a los diferentes ecosistemas donde pueden cultivarse (sur y este de la península en la costa mediterránea y Canarias), especialmente en lo que se refiere a su adaptabilidad a los suelos calcáreos y al empleo de aguas salinas, bicarbonatadas o cloruradas que se presentan en el área de Canarias. Otro problema a resaltar en estos cultivos es el de los patógenos (hongos) que provocan podredumbres en raíces y el de las virosis y viroides (sunblotch). Finalmente, cabe señalar como problema también grave, tanto del aguacate como del chirimoyo, una irregular fructificación, irregularidad cuyas causas se nos escapan, debido sin duda a nuestros escasos conocimientos sobre la fisiología de su fructificación.

Muchos y muy diversos son los problemas a los que se enfrenta el cultivo de las hortalizas de exportación. Empecemos por el de la escasa calidad de las semillas de las diferentes especies y variedades de hortalizas. En general, en España se recurre al empleo de semillas de importación para cultivar hortalizas exportables, ante la menor productividad de las variedades autóctonas o en razón de que éstas producen frutos que no satisfacen las necesidades del consumidor de los países importadores. Para mantener el prestigio en el mercado hay que concurrir con productos de la máxima calidad, y puesto que constan-

temente se está mejorando tanto el aspecto externo (presentación del producto) como sus cualidades culinarias y nutritivas, el horticultor español ha de recurrir al empleo de semillas que, generalmente, encuentra en otros países y no en el nuestro, por carecer de la infraestructura necesaria en mejora vegetal para satisfacer la demanda de productos de calidad. Se ha de ir más allá de la simple mejora de la calidad de los frutos, y desarrollar variedades de calidad que sean, al mismo tiempo, resistentes a plagas y enfermedades endémicas en el país.

No menos grave es el problema del cansancio del suelo. La proliferación de la población de nemátodos, hongos y bacterias del suelo, consecuencia del cultivo prolongado de determinadas hortalizas, incide fuertemente en el rendimiento de las cosechas. Este problema, que antaño se combatía con la rotación de los cultivos de forma poco satisfactoria, se combate hogaño mediante los tratamientos de desinfección del suelo que, si bien son eficaces, resultan muy costosos y merman considerablemente la rentabilidad de la cosecha. No se ha prestado la atención debida al problema que representan las fitotoxinas en el suelo, aunque se está ya profundizando en su conocimiento, origen, producción e incidencia sobre la horticultura y sobre los métodos de lucha contra las mismas. La falta de conocimientos precisos y concretos sobre los agentes o circunstancias que causan el cansancio del suelo incide de forma negativa en la rentabilidad de las cosechas de hortalizas. Otro problema delicado es el relativo a las pérdidas en el almacenamiento post-recolección de los productos hortícolas. Los productos que se pierden anualmente después de cosechados, como consecuencia del ataque de insectos y hongos, constituyen un grave problema, para cuya solución habrá que atender desde el acondicionamiento (en todas sus acepciones) de los locales y depósitos de almacenamiento hasta los productos y tipos de tratamientos que deben aplicarse. Los roedores constituyen una plaga grave en algunas zonas del país, por los daños que ocasionan en los productos agrícolas almacenados. Al considerar este problema es menester tener en cuenta, también, la incidencia que tanto algunos de los hongos causantes de pérdidas (*Aspergillus* productores de aflatoxinas) como los productos utilizados para combatirlos (fungicidas, fumigantes, etcétera) y los residuos de los mismos tienen sobre la salud del consumidor. La corrección del problema presenta, pues, una doble faceta: por un lado, se han de evitar las cuantiosas pérdidas que se originan anualmente y, por otro, se ha de garantizar la salubridad de los productos.

Veamos, por último, el capítulo de flores y plantas ornamentales. De nuevo, uno de los problemas más importantes a los que se enfrenta nuestra floricultura estriba en la falta de conocimientos sobre las condiciones de cultivo más adecuadas tanto para las especies del mercado internacional como para las especies autóctonas que se prevé pueden competir con ventaja en aquel mercado. La deficiente estructura productiva y de los canales de comercialización son los más acusados con los que se enfrenta el cultivo de ornamentales en España que, por sus condiciones geográficas y climatológicas, puede producir una gama muy amplia de artículos que gozan de gran demanda. Se requiere un conocimiento preciso de las exigencias de cultivo de estas especies, la selección de las mismas y su adaptación a las condiciones ecológicas. Por otro lado, la floricultura española puede ofrecer artículos de difícil producción en los países del centro y norte de Europa. La situación de crisis energética por la que pasa el mundo es otro factor que nos permite competir con ventaja con los países productores de flores.

No conviene pasar de largo por la deficiente condición sanitaria del material vegetal cultivado actualmente; la expansión del cultivo de bulbos productores de flores (tulipanes, jacintos, gladiolos, etcétera) se ve frenada por dificultades sanitarias. Conviene establecer, cuanto antes, programas de mejora sanitaria en este sector.

Tanto el abonado como el riego y los tratamientos fitosanitarios se están llevando a cabo sin el adecuado soporte técnico. En algunas zonas, como Almería y Málaga, donde se practica el cultivo enarenado, es necesario profundizar en la peculiaridad de esta técnica en lo referente a las diferentes operaciones del cultivo (fertilización, riego, tratamientos fitosanitarios) para informar a los agricultores y proceder a las correcciones oportunas.

Deberíamos detenernos, quizás, en los tipos de protección utilizados para forzar el cultivo y el manejo de los mismos, en la fisiología post-recolección de los diferentes productos y en otros problemas no menos graves de nuestra hortofruticultura. Pero creemos que con los hasta aquí reseñados podemos tener una panorámica completa de la situación actual de la horticultura en España. (P. Cuñat, A. Aguilar y V. García.)

La galaxia de la Vía Láctea

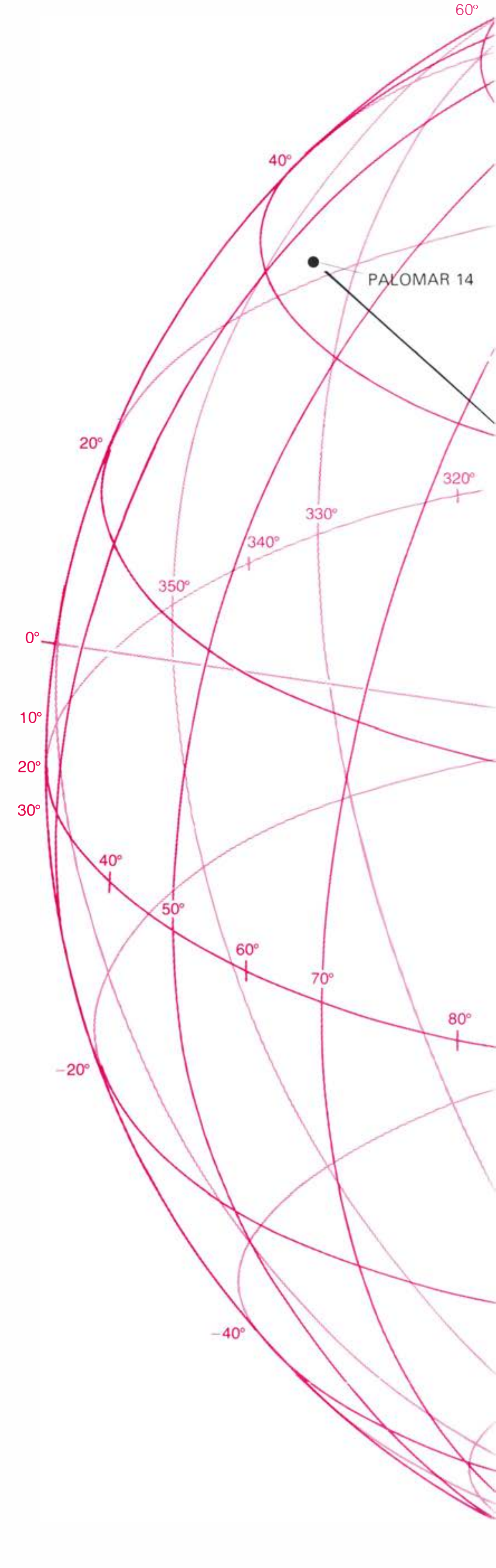
Hace algunos años, los datos fundamentales de la galaxia parecían definitivamente establecidos. Pero hoy se cuestiona hasta su masa y su radio

Bart J. Bok

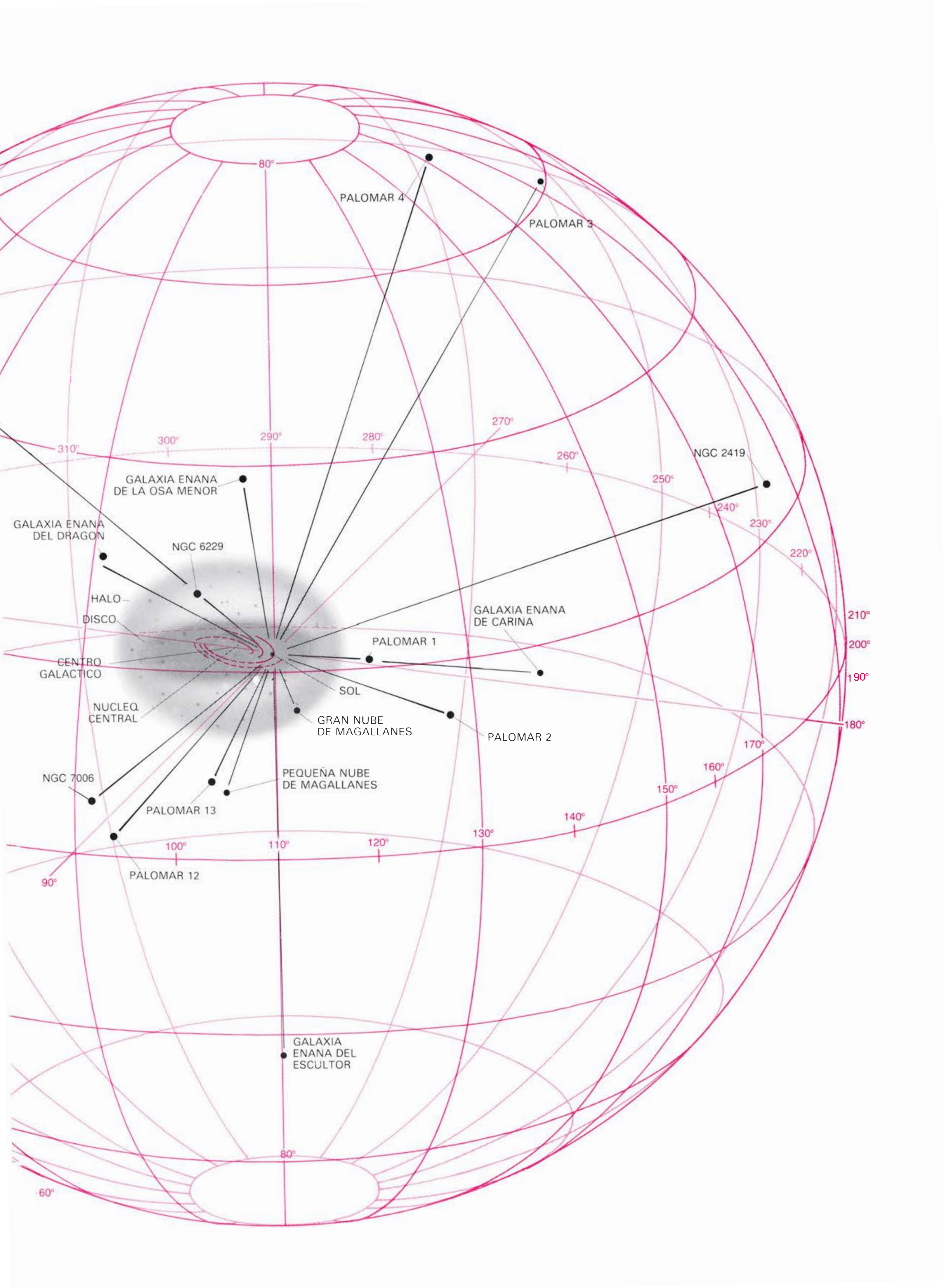
En una noche clara y sin luna, libre de las luces de la civilización, no hay cosa en el cielo que nos impresione tanto como la banda luminosa de la Vía Láctea. Incluso sin telescopio, mucho es lo que puede decirse sobre ella. Se puede ver, por ejemplo, que las estrellas se hacen más numerosas cuando la mirada recorre el cielo hacia la banda. El hecho de que ésta conste principalmente de estrellas nos parece entonces menos sorprendente. Se puede ver que la banda sigue un círculo máximo que divide en dos la esfera celeste; en consecuencia, la Tierra está inmersa en el plano central de la banda. Se puede ver, además, que presenta su máxima amplitud y brillo en dirección de la constelación de Sagitario. Ciertamente esta es la dirección que conduce hacia el centro del sistema. El sistema es la galaxia de la Vía Láctea.

Un telescopio revelará que algunos puntos de la Vía Láctea semejantes a estrellas son, en realidad, grandes agrupaciones de estrellas. Se trata claramente de objetos distantes, que se conocen como cúmulos globulares. Entre 1918 y 1921, Harlow Shapley empleó los telescopios del Observatorio del Monte Wilson para demostrar que los cúmulos, lo mismo que las estrellas en la banda, presentan su máxima frecuencia en Sagitario. En una región de Sagitario que constituye sólo el 2 por ciento del cielo, Shapley inscribió una tercera parte de todos los cúmulos globulares entonces conocidos. Resulta, pues, evidente que el sistema solar está lejos del centro de la galaxia; la distancia del Sol al centro se estima ahora en 8500 parsecs. (Un parsec equivale a 3,26 años-luz.) Desde entonces se ha establecido que el sistema solar, junto con el resto de la masa del plano central de la galaxia, gira en torno al centro. El Sol gira a una velocidad que ahora se calcula en unos 230 kilómetros por segundo. Por tanto, completa una revolución cada 200 millones de años.

En 1930, Robert J. Trumpler, del Observatorio Lick, mostró que un medio interestelar de gas y polvo atenúa la luz de las estrellas, particularmente las que están en el plano central de la galaxia. En 1951, William W. Morgan, del Observatorio Yerkes, y sus alumnos Donald E. Osterbrock y Stewart L. Sharpless hallaron indicaciones de que la banda no era otra cosa que la estructura espiral de la galaxia vista "de canto" desde la Tierra. La investigación en la banda de ondas de radio pronto produjo pruebas de las características espirales. En el decenio de 1960, Chia



MAPA DE LA VIA LACTEA donde se muestra la galaxia de acuerdo con la hipótesis según la cual posee un tamaño inesperadamente grande y es masiva. Los tres componentes de la galaxia conocidos de antiguo encajan perfectamente dentro del globo coordenado. Las dimensiones se dan en parsecs. (Un parsec equivale a 3,26 años-luz.) Entre las tres componentes, la porción central, con un radio de 4000 a 5000 parsecs, consta principalmente de un denso amontonamiento de estrellas antiguas. El disco galáctico, cuyo radio mide 15.000 parsecs, consta de estrellas más jóvenes y polvo y gas. Sus configuraciones espirales (*curvas de color*) sólo se han podido reconocer en la vecindad del Sol. El halo galáctico, con un radio de 20.000 parsecs, consta, principalmente, de un delgado agrupamiento de estrellas antiguas y aproximadamente la mitad de las agregaciones estelares llamadas cúmulos globulares. El hipotético componente más externo de la galaxia se llama corona; su presencia se ha deducido de las velocidades de la materia visible. Presumiblemente, los objetos de la corona no destacan por su brillo. En esta figura, los "compañeros galácticos" visibles en la corona hasta una distancia de 100.000 parsecs se representan en coordenadas centradas en el Sol. La dirección del Sol al centro de la galaxia define la longitud galáctica de cero grados. Los ángulos por encima y por debajo del plano galáctico se miden en latitud galáctica. Los compañeros incluyen 10 cúmulos globulares, cuatro galaxias esferoidales enanas y las galaxias irregulares en la Gran Nube y Pequeña Nube de Magallanes. La distancia del Sol al centro ronda los 8500 parsecs.



80°
PALOMAR 4
PALOMAR 3
270°
310°
300°
290°
280°
260°
NGC 2419
250°
240°
230°
220°
GALAXIA ENANA
DE LA OSA MENOR
GALAXIA ENANA
DEL DRAGON
NGC 6229
HALO
DISCO
GALAXIA ENANA
DE CARINA
210°
200°
PALOMAR 1
CENTRO
GALACTICO
190°
SOL
NUCLEO
CENTRAL
GRAN NUBE
DE MAGALLANES
180°
PALOMAR 2
170°
NGC 7006
PEQUEÑA NUBE
DE MAGALLANES
160°
150°
PALOMAR 13
140°
130°
100°
110°
120°
90°
PALOMAR 12
GALAXIA
ENANA DEL
ESCULTOR
80°
60°

Chiao Lin, del Instituto de Tecnología de Massachusetts, y Frank H. Shu, entonces en el Observatorio del Harvard College, sugirieron que las ondas de densidad creciente en el medio interestelar producían los rasgos espirales y eran responsables de la formación de estrellas. Mientras tanto, el advenimiento de la astronomía de infrarrojo proporcionó un medio de explorar el interior de las oscuras nubes interestelares. Parece ahora que las nubes están donde se forman la mayor parte de las nuevas estrellas de la galaxia. Hacia 1970, la radioastronomía había comenzado a revelar la composición de las nubes oscuras. Están formadas de polvo e hidrógeno, con una mezcla de algunas moléculas sorprendentemente complicadas, algunas de ellas orgánicas.

Durante muchos años, he sido un asiduo observador nocturno de la galaxia de la Vía Láctea. Recuerdo la época de mediados del decenio de 1970 como aquella en la que mis compañeros de observación y yo estábamos notablemente confiados. Las características principales de la galaxia parecían razonablemente bien establecidas. La galaxia tenía dos componentes principales: un denso núcleo central de estrellas, con el límite entre 4000 y 5000 parsecs del centro, y un disco plano y mucho más delgado de estrellas y gas y polvo interestelar, cuyo borde interior lindaba con el núcleo y cuyo borde exterior distaba unos 15.000 parsecs de él. El núcleo está en Sagitario; el disco se extiende a través del cielo.

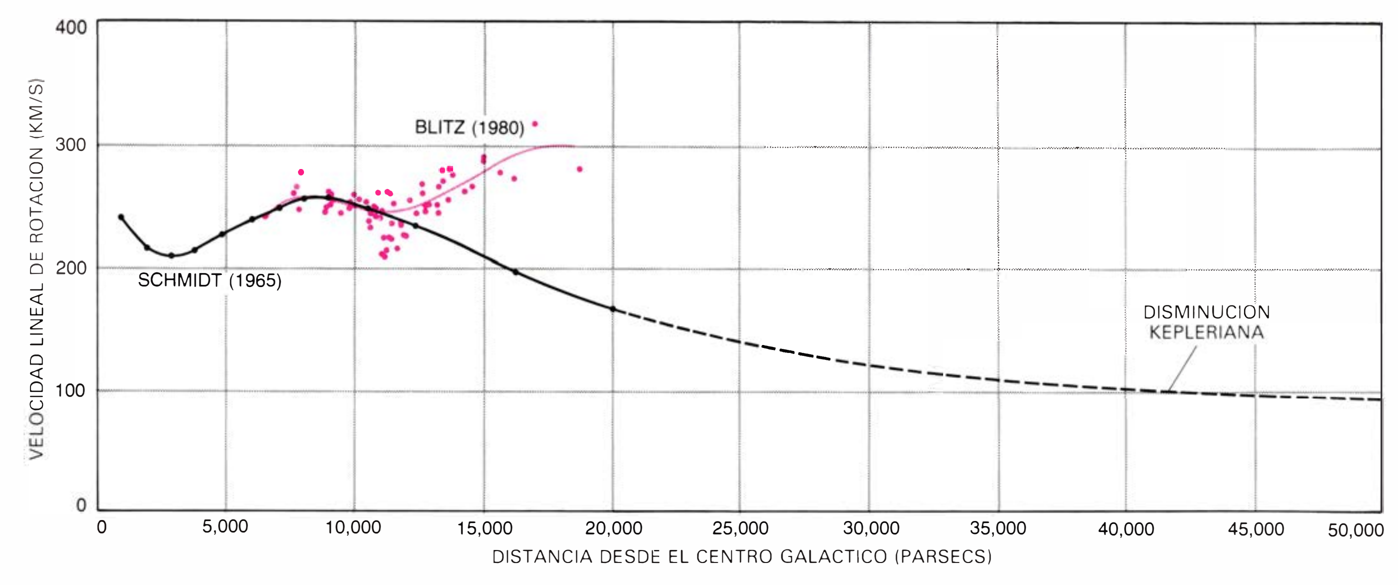
La masa combinada del núcleo y el disco se calculaba entonces que era inferior a 200.000 millones de veces la masa solar. El disco y el núcleo, sin embargo, estaban rodeados de un "halo" de materia a cada lado del plano central de la galaxia. En términos generales, el halo tiene la forma de una esfera ligeramente achatada. En el plano de la galaxia, su radio mide unos 20.000 parsecs. Es notable por sus estrellas antiguas y también por una dispersión de unos 100 cúmulos globulares. (Otros 100 cúmulos globulares están en el disco galáctico, o en su vecindad.) El halo podría añadir a lo más 100.000 millones de masas solares a la masa total de la galaxia.

Quedaba todavía mucho trabajo. Pero confiábamos en que los hechos básicos se mantendrían y que los astrónomos no se llevarían mayores sorpresas cuando investigaran cuestiones tales como la formación de las estrellas y la manera en que las características espirales de la Vía Láctea evolucionan. No sospechábamos que pronto sería necesario revisar el radio de la Vía Láctea multiplicándolo por un factor de tres o más y aumentar su masa multiplicándola nada menos que por un factor 10. Las revisiones son representativas de cierto número de recientes cataclismos. Consideraré aquí varios aspectos del trabajo empeñado hoy en la comprensión de la Vía Láctea. Me he propuesto abarcar la mayoría de las líneas de investigación porque se trata de áreas de notable fermento, donde el avance de nuestro conocimiento puede darse de un momento a otro. A modo de contraste, incluyo además una línea particular de la que temo que esté llegando a un callejón sin salida.

Las coronas galácticas

Ya en 1974 se ofrecieron indicaciones de que la Vía Láctea era, contra todo pronóstico, grande y masiva. Las sugerencias venían, principalmente, del terreno teórico, avaladas por Donald Lynden-Bell, de la Universidad de Cambridge, y de Jeremiah P. Ostriker, P. J. E. Peebles y Amos Yahil, de la Universidad de Princeton. El argumento básico era que no se podían garantizar la estabilidad dinámica y la permanencia de la galaxia a menos que el disco galáctico estuviera rodeado y, por consiguiente, gravitatoriamente estabilizado, a pesar de su delgadez y delicada estructura espiral, por un halo extenso y masivo.

J. Einasto y colaboradores, del Observatorio de Tartu, en Estonia, presentaron un razonamiento diferente. Durante varios años, los investigadores ocupados en la dinámica de la Vía Láctea habían argumentado que la velocidad del Sol respecto a los cúmulos globulares en el halo de la galaxia era de sólo unos 180 kilómetros por segundo. Como los cúmulos estaban dispersos en un gran volumen esférico, parecía claro que globalmente no participaran en cualquier rotación que el disco galáctico pudiera poseer. La distribución de los cúmulos, por tanto, habría de ser



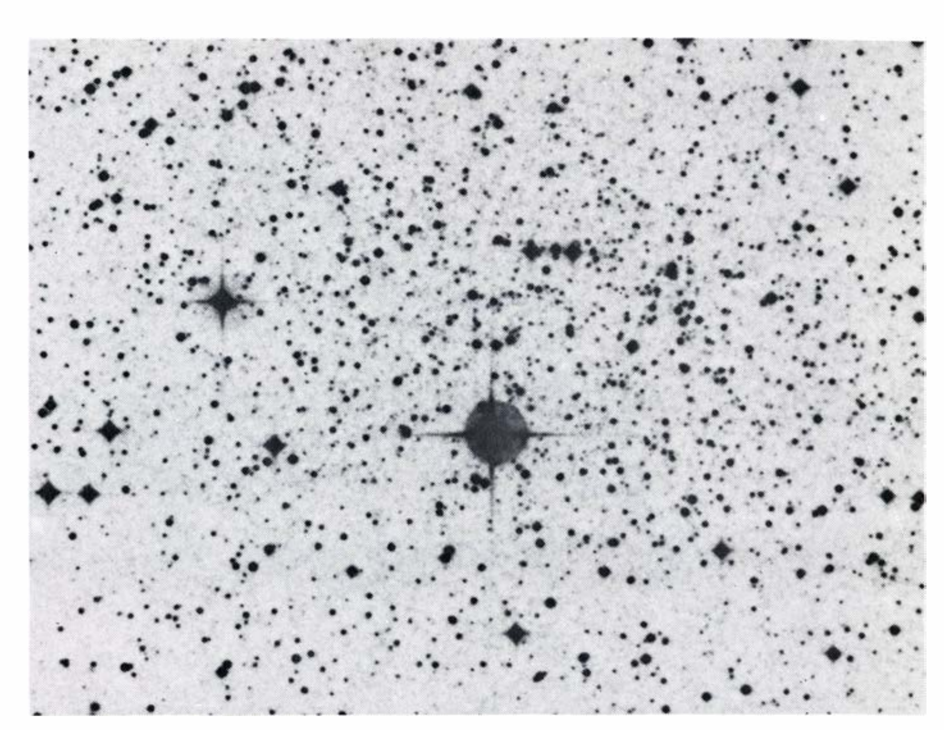
LA CURVA DE ROTACION representa gráficamente la velocidad lineal de la materia en rotación alrededor del centro de la Vía Láctea. Aquí se han construido dos de tales curvas. Una curva de rotación, trazada en 1965 por Maarten Schmidt, de los Observatorios Hale (*curva negra continua*), muestra una velocidad lineal que disminuye hacia el límite de la galaxia visible, a 20.000 parsecs. Si toda la masa de la galaxia estuviera comprendida dentro de tal límite, una masa de prueba colocada más allá giraría a una velocidad menor, obedeciendo con bastante aproximación una ley que formulara Johannes Kepler para el movimiento de los planetas (*línea de trazos*). Datos analizados por Leo Blitz y sus colegas, de Berkeley, sirven para dibujar una curva de rotación (*línea de color*) que crece hasta 300 km por segundo, a 20.000 parsecs. El máximo de la nueva curva implica la existencia de una masa invisible fuera, en parte, del límite visible de la galaxia. Cada punto de los datos más recientes simboliza la velocidad lineal de una nube de átamos de hidrógeno.

más o menos estacionaria con respecto al centro galáctico. Específicamente, se sostiene que la velocidad de rotación del sistema de cúmulos globulares en torno al centro de la Vía Láctea quizá no supere los 50 kilómetros por segundo. Por lo tanto, la velocidad de rotación del Sol alrededor de la galaxia no debe ser superior a 230 kilómetros por segundo, y ciertamente no superior a los 250.

Mientras tanto, otros investigadores habían determinado la velocidad del Sol con respecto al movimiento medio de las galaxias vecinas, es decir, aquellas cuya distancia del Sol no es mayor que un millón de parsecs. Su resultado para la velocidad de rotación del Sol alrededor del centro de la Vía Láctea era de 300 kilómetros por segundo. La diferencia entre los dos resultados podía interpretarse de la manera más verosímil como una indicación de que el centro de la Vía Láctea se estaba moviendo a una velocidad de 50 a 80 kilómetros por segundo respecto de las galaxias vecinas. Esa velocidad parecía sorprendentemente grande.

Einasto argumentaba que, pese a ello, el resultado podría ser correcto y que su causa podría ser la masa insospechadamente grande del sistema de la Vía Láctea. Para poner a prueba su hipótesis, examinó el movimiento de la Vía Láctea dentro del grupo de galaxias al que pertenece. Einasto propuso la existencia de subgrupos dentro del grupo, y que, en uno de los subgrupos, la Vía Láctea dominaba gravitatoriamente sobre varias otras agregaciones de estrellas, tales como las dos pequeñas galaxias próximas, llamadas Pequeña Nube y Gran Nube de Magallanes y un miembro de las galaxias esferoidales enanas de las que en la actualidad se conocen siete. Una de las galaxias esferoidales enanas se encuentra a unos 150.000 parsecs del centro de la Vía Láctea. La velocidad del Sol respecto al movimiento medio de esos acompañantes galácticos demostró ser de casi 300 kilómetros por segundo. Volvía a ser sorprendentemente grande. Einasto interpretó la velocidad como un efecto de la gran masa de nuestra galaxia sobre las galaxias vecinas más próximas.

Por tanto, se reforzó la sospecha de Einasto de que la Vía Láctea fuera más extensa y más masiva de lo que se había supuesto. En 1976 presentó un modelo del sistema de la Vía Láctea donde la masa del núcleo central, el disco y un halo extenso era de 900.000 millones de masas solares. Pero ese valor seguía siendo insuficiente para explicar las grandes diferencias de velocidad obser-



GALAXIA ENANA CARINA, última adición al catálogo de compañeros galácticos; su distancia, estimada en unos 100.000 parsecs del centro de la Vía Láctea, la convierte, según todos los indicios, en un miembro externo de la corona galáctica. En este negativo fotográfico, la galaxia enana es la agregación suelta y apenas destacada de pequeños puntos negros cerca del centro. Los puntos mayores corresponden a estrellas mucho más próximas al sistema solar situadas a lo largo de la misma visual. La fotografía fue realizada como parte del programa de exploración del Observatorio Meridional Británico-Europeo.

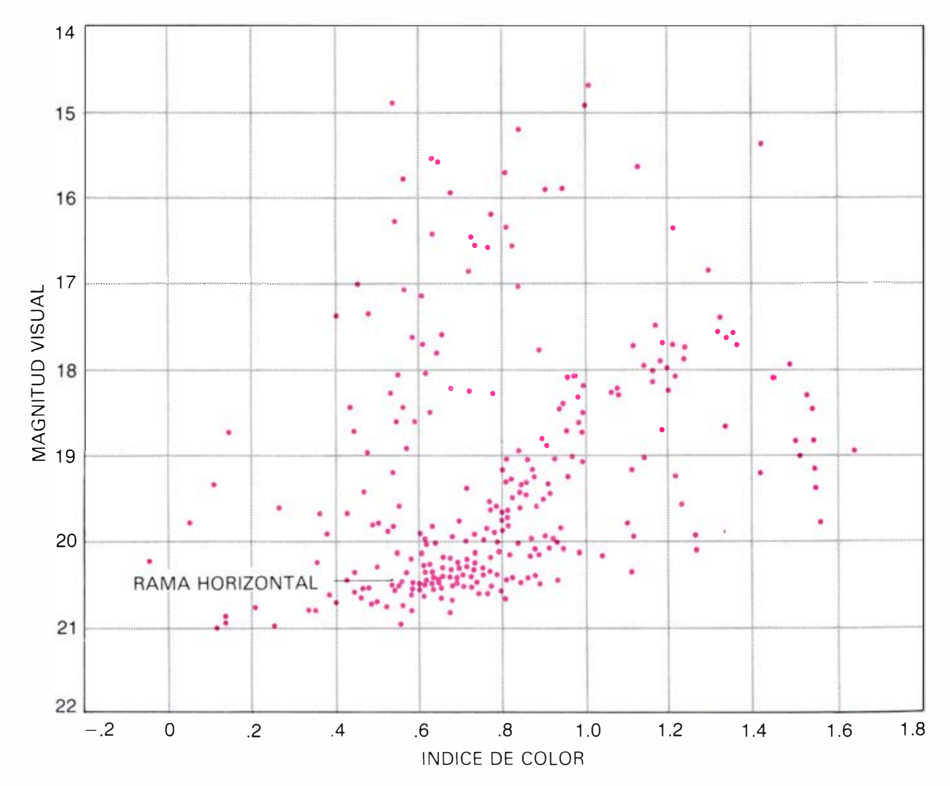
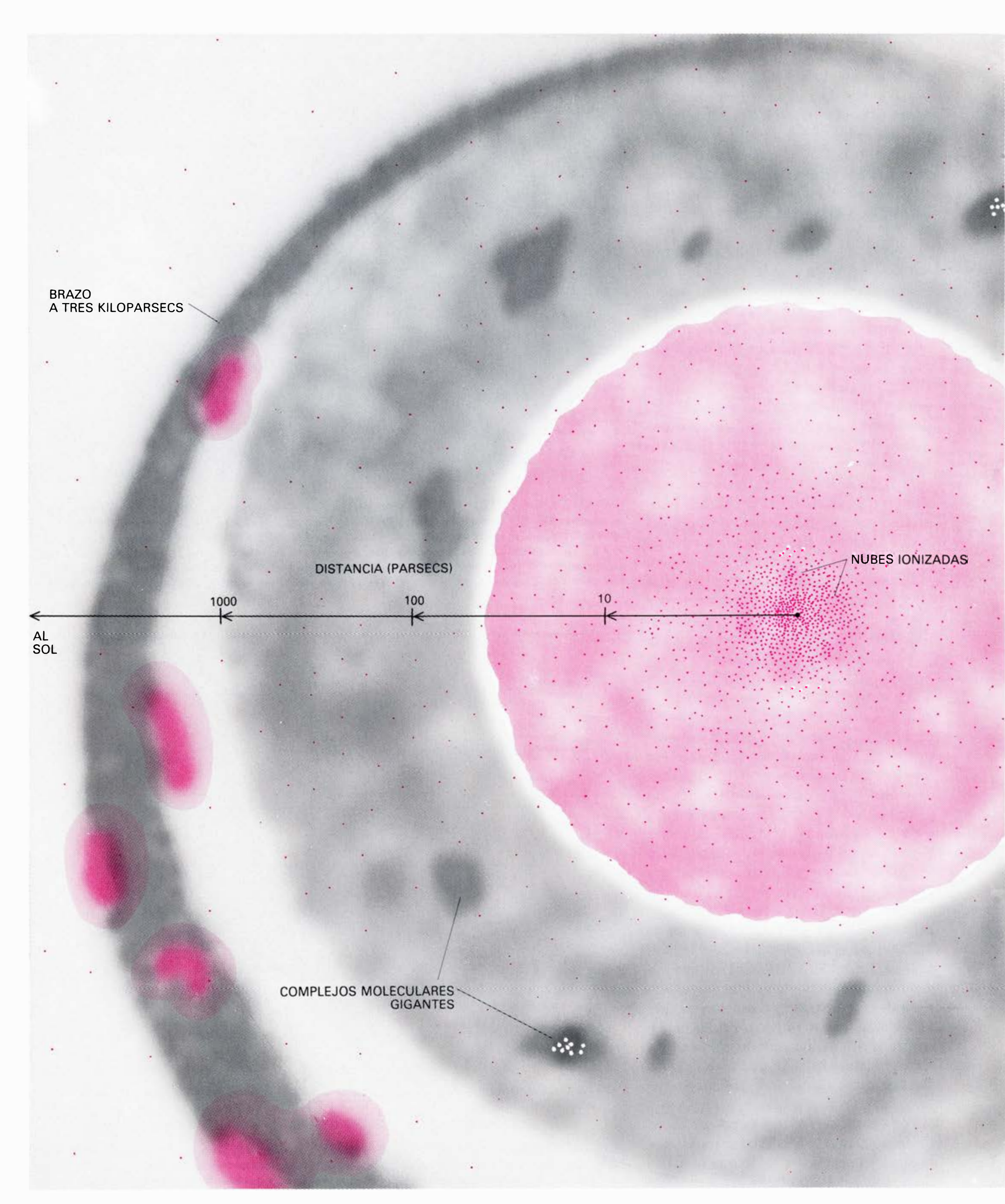
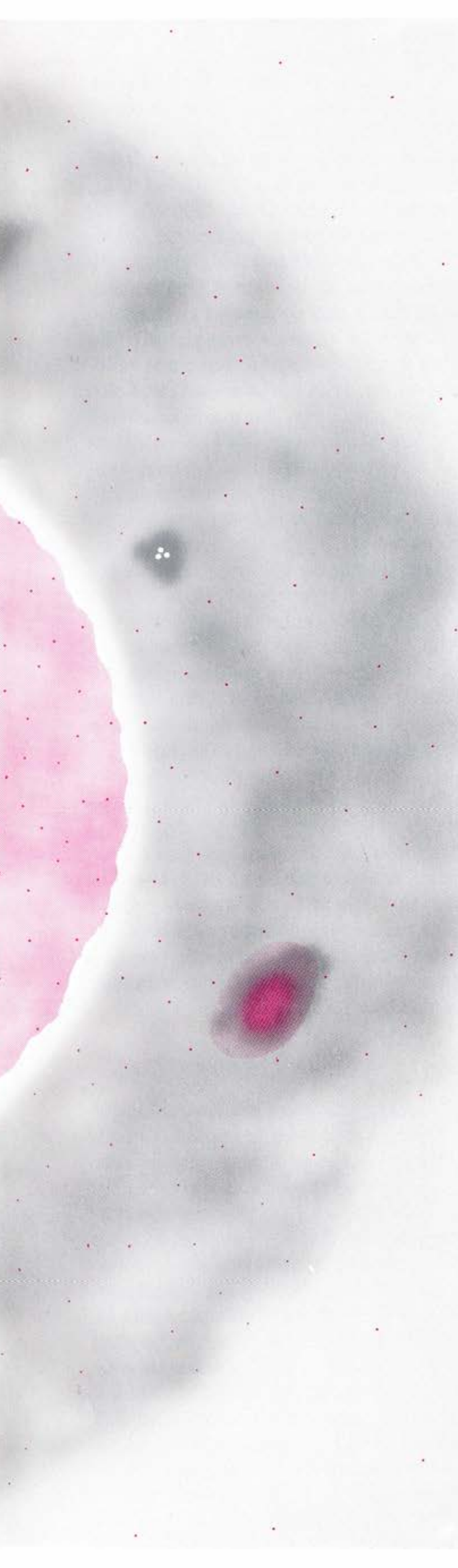


DIAGRAMA DE HERTZSPRUNG-RUSSELL para las estrellas en la parte central de la galaxia enana de Carina. El diagrama permite determinar la distancia de ese compañero galáctico. Representa la luminosidad aparente de cada estella (*eje vertical*) en función de una medida de su color (*eje horizontal*). Las estrellas en lo que se llama la rama horizontal del diagrama resultan tener una luminosidad aparente de aproximadamente 20,4. Su luminosidad intrínseca las colocaría 20 magnitudes más arriba. A partir de la atenuación de su luminosidad, se deduce la distancia, cifrada en 100.000 parsecs. A la latitud galáctica de − 22 grados, la galaxia enana de Carina se halla fuera del plano central de la Vía Láctea; de aquí que la corrección del cálculo de la absorción de la luz por la materia interestelar interpuesta deba ser pequeño. Los datos fueron recogidos por Russell Cannon en el Observatorio Real de Edimburgo y en el Observatorio Anglo-Australiano. El límite de sensibilidad del aparato usado por los investigadores queda justamente por debajo de la rama horizontal. (En el mapa de la Vía Láctea puede apreciarse Carina a la derecha del Sol.)



PARTE CENTRAL DE LA VIA LACTEA, esquematizada en esta ilustración ideada por Thomas R. Geballe, de los Observatorios Hale. La galaxia está vista desde arriba. La escala decrece logarítmicamente con la distancia al centro, de modo que los parsecs más cercanos a éste aparecen aumentados. Los tres parsecs más próximos al centro incluyen la congestión más densa de estrellas (*punto de color*) en la galaxia. (En toda la ilustración, hemos dividido la densidad de estrellas por 2000.) Esta región aloja también compactas nubes de gas ionizado (*color oscuro*) constituidas principalmente por hidrógeno. La velocidad de las nubes parece indicar que giran en torno a un objeto de masa extraordinariamente grande, un agujero negro tal vez, situado precisamente en el centro. Una distribución más tenue de gas ionizado (*color claro*) impregna los 100 parsecs centrales. Está rodeada a su vez por un anillo de gas más frío y no ionizado (*gris*) cuyo hidrógeno se presenta en su forma atómica y en su forma molecular. El anillo incluye complejos gigantes de polvo y moléculas



(*gris oscuro*). En algunos de ellos se han originado estrellas jóvenes (*puntos blancos*). A la izquierda aparece una densa banda constituida principalmente por hidrógeno no ionizado. Es la parte más interna de una estructura en expansión llamada el brazo a tres kiloparsecs. El brazo pudo originarse por una explosión en el centro de la galaxia.

vadas entre esta galaxia y sus compañeras. Por consiguiente, Einasto propuso que el núcleo, el disco y el halo se hallaban inmersos en una componente, todavía mayor y sin embargo invisible, de la galaxia, a la que llamó corona, que se extendía hasta al menos 100.000 parsecs del centro y tenía una masa de 1,2 billones de masas solares. La masa total de la Vía Láctea sería entonces de 2,1 billones de masas solares, es decir, al menos siete veces el valor aceptado en 1975.

Pronto llegaron indicaciones a favor de esta hipótesis, suministradas por distintos investigadores. En primer lugar, Vera C. Rubin, W. Kent Ford, Jr., y Norbert Thonnard, del Departamento de Magnetismo Terrestre de la Institución Carnegie de Washington, examinaron los corrimientos Doppler de líneas espectrales en la luz emitida por la materia en la región periférica de 17 galaxias. Cada uno de tales corrimientos es una traslación de una línea espectral a una longitud de onda diferente de la que tendría si la fuente de radiación estuviera inmóvil respecto al instrumento que la recibe. Los investigadores llegaron a la conclusión de que la materia cortical en cada galaxia giraba, alrededor del centro de ésta, por lo menos tan deprisa como la materia más próxima al centro. En otras palabras, la curva de rotación –o gráfica de la velocidad de rotación en función de la distancia al centro de la galaxia– era esencialmente una línea horizontal para la parte externa de cada una de las galaxias que estudiaron.

Debe destacarse, por su importancia, el descubrimiento de tales curvas de rotación lineales. Después de todo, la distribución del brillo en la imagen óptica de una galaxia espiral típica conduce a inferir que la materia visible de la galaxia está concentrada hacia el centro y se hace más tenue hacia la periferia. De esta concentración se puede deducir que la materia visible más externa debería estar moviéndose en respuesta a fuerzas análogas a las que actúan sobre los planetas más externos del sistema solar. La materia periférica debería estar girando alrededor del centro de la galaxia a una velocidad lineal menor (expresada en kilómetros por segundo) que la de la materia más próxima al centro. Evidentemente, no es así. Las curvas de rotación calculadas por Rubin, Ford y Thonnard implican, pues, la presencia de materia invisible en gran cantidad más allá de la periferia aparente de cada galaxia.

Estudios análogos a los de Rubin, Ford y Thonnard se centran directamente en la distribución de masa en la Vía Láctea. Así, F. D. A. Hartwick, de la Universidad de Victoria, y Wallace L. W. Sargent, del Instituto de Tecnología de California, calcularon las velocidades de los cúmulos globulares a distancias superiores a 20.000 parsecs del centro de la galaxia. En otros estudios, James E. Gunn, Gillian R. Knapp y Scott D. Tremaine emplearon datos recogidos en el Observatorio Radioastronómico de Owens Valley para determinar las velocidades de las nubes de átomos de hidrógeno interestelares. Maurice P. Fitzgerald, de la Universidad de Waterloo, junto con Peter D. Jackson, de la Universidad de Maryland y Anthony Moffat, de la Universidad de Montreal, determinaron las velocidades de estrellas y cúmulos de estrellas a distancias hasta 17.000 parsecs del centro de la galaxia. William L. H. Shuter, de la Universidad de Columbia Británica, dedujo las velocidades de nubes de hidrógeno y monóxido de carbono situadas en el anticentro galáctico, o dirección opuesta a la visual desde la Tierra al centro de la galaxia. Recientemente, Leo Blitz y sus colegas, de la Universidad de California en Berkeley, han medido el corrimiento Doppler en las regiones ópticas y de radio del espectro electromagnético para líneas procedentes de 184 nebulosas y grandes nubes de hidrógeno interestelar y monóxido de carbono en el anticentro. Todos los resultados concuerdan en dar una curva de rotación que no desciende.

La corona de la Vía Láctea

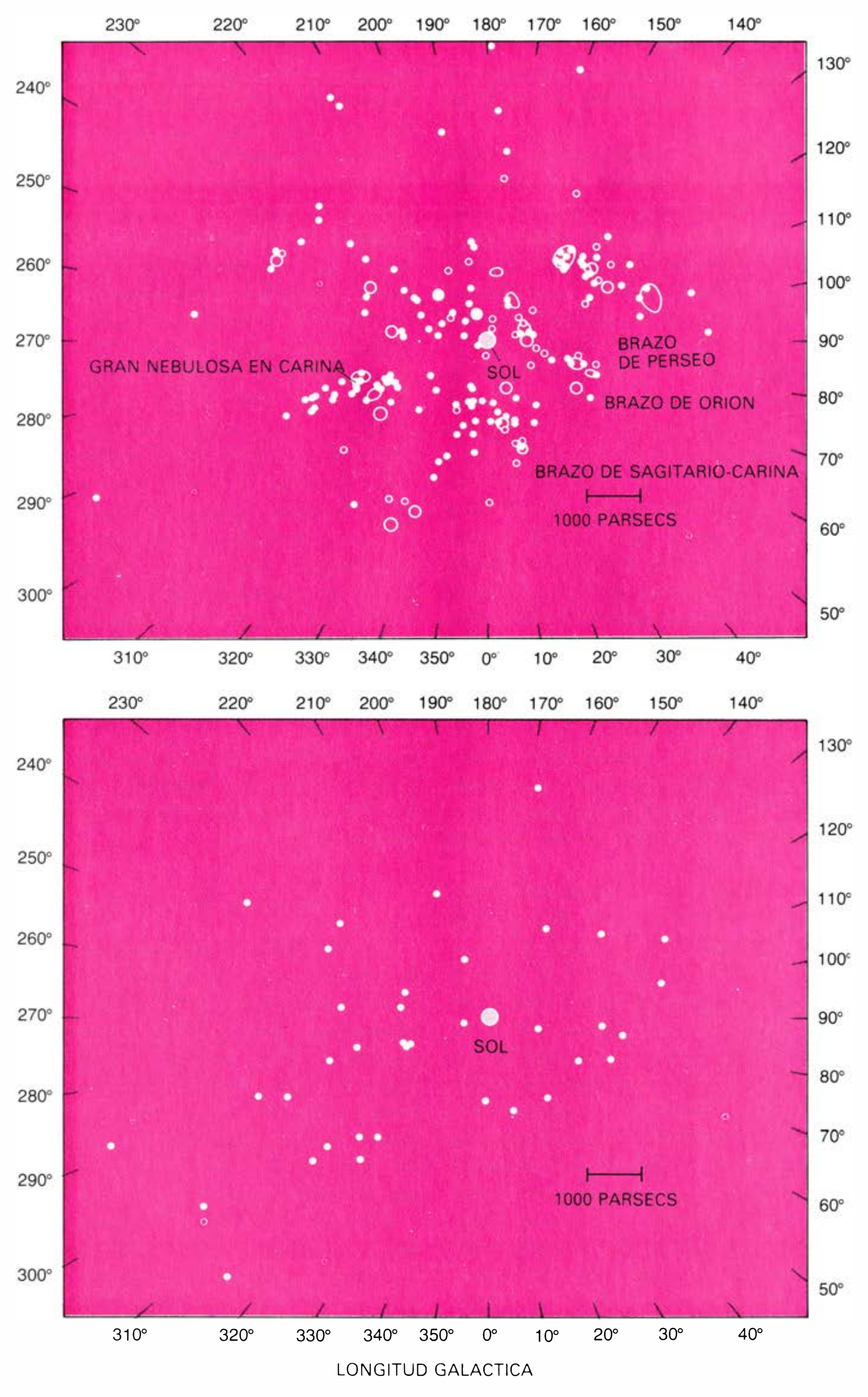
Con los datos disponibles hoy, parece natural llegar a la conclusión de que la curva de rotación para la Vía Láctea alcanza un valor de 230 kilómetros por segundo a 8500 parsecs, la distancia desde el centro galáctico que marca la posición del Sol. Desde allí, la velocidad de rotación continúa aumentando. Alcanza 300 kilómetros por segundo a una distancia de 20.000 parsecs.

Existen varios objetos visibles a distancias superiores a los 20.000 parsecs: la Unión Astronómica Internacional publicó una lista de ellos en 1979. Hay cuatro cúmulos globulares entre 20.000 y 40.000 parsecs del centro galáctico; la Gran Nube de Magallanes y dos cúmulos globulares se encuentran entre 40.000 y 60.000 parsecs; dos galaxias esferoidales enanas y la Pequeña Nube de Magallanes se hallan entre 60.000 y 80.000 parsecs; una galaxia esferoidal enana y tres cúmulos globulares se localizan entre 80.000 y 100.000 parsecs.

Cuatro galaxias esferoidales enanas más y dos cúmulos globulares se encuentran a distancias entre 100.000 y 220.000 parsecs, pero sus pretensiones de pertenecer a la corona de la Vía Láctea son más dudosas.

Claramente, nuestra galaxia se extiende mucho más y tiene una masa mucho mayor de lo que hasta aquí se pensaba; la Vía Láctea ha sido ascendida a la categoría de gran galaxia espiral. Tomados juntos, sin embargo, los componentes visibles de la corona tienen sólo una pequeña fracción de los cientos de miles de millones de masas solares que deberían estar allí. Al parecer, la Vía Láctea comparte con las galaxias estudiadas por Rubin, Ford y Thonnard la propiedad de que una fracción notable de la materia situada al exterior es oscura. Y así, John H. Bahcall y Raymond M. Soneira, del Instituto de Estudios Superiores, deducen de la invisibilidad de la materia hipotética que, si hay estrellas en la corona que no están agrupadas en cúmulos o dentro de galaxias enanas, su brillo intrínseco debe ser inferior a una milésima parte de la luminosidad del Sol.

¿Qué es, pues, la masa invisible? Hay tres hechos que vale la pena destacar. En primer lugar, las galaxias enanas y los cúmulos globulares están constituidos, en su mayor parte, por estrellas antiguas. En segundo lugar, las estrellas antiguas no son altamente luminosas. En tercer lugar, nadie ha detectado, procedentes de la corona, las líneas espectrales que caracterizan las nubes de materia gaseosa, tales como hidrógeno y monóxido de carbono, que hay en lugares más hacia el centro de la galaxia. En el momento presente, por tanto, la hipótesis más fundada es aquella que afirma que la corona de la Vía Láctea está formada principalmente por estrellas antiguas que han agotado su combustible. Por otro lado, la masa invisible de la corona de la galaxia quizá no encaje en ninguna de las categorías basadas en lo que puede verse en regiones más accesibles. No sabemos todavía qué hay allí fuera.



ESTRUCTURA ESPIRAL del disco de la Vía Láctea en la vecindad del Sol; se ha investigado por observaciones en longitudes de onda del visible con éxito variable, dependiendo de las clases de objetos astronómicos empleados en el intento. La figura superior muestra las posiciones de grupos sueltos de estrellas jóvenes (*circunferencias*) y de cúmulos en los que al menos algunas de las estrellas son jóvenes (*círculos llenos*); la distribución de tales estrellas confirma la presencia de las estructuras espirales llamadas brazo de Perseo, brazo de Orión y brazo de Sagitario-Carina. La figura inferior recoge las posiciones de las Cefeidas variables de largo período. Estas jóvenes estrellas gigantes se forman, según se cree, dentro de los segmentos espirales. Aunque se representa la misma parte de la galaxia, no se observa estructura espiral. Los datos para las figuras fueron recogidos por Roberta M. Humphreys, de Minnesota.

El núcleo central

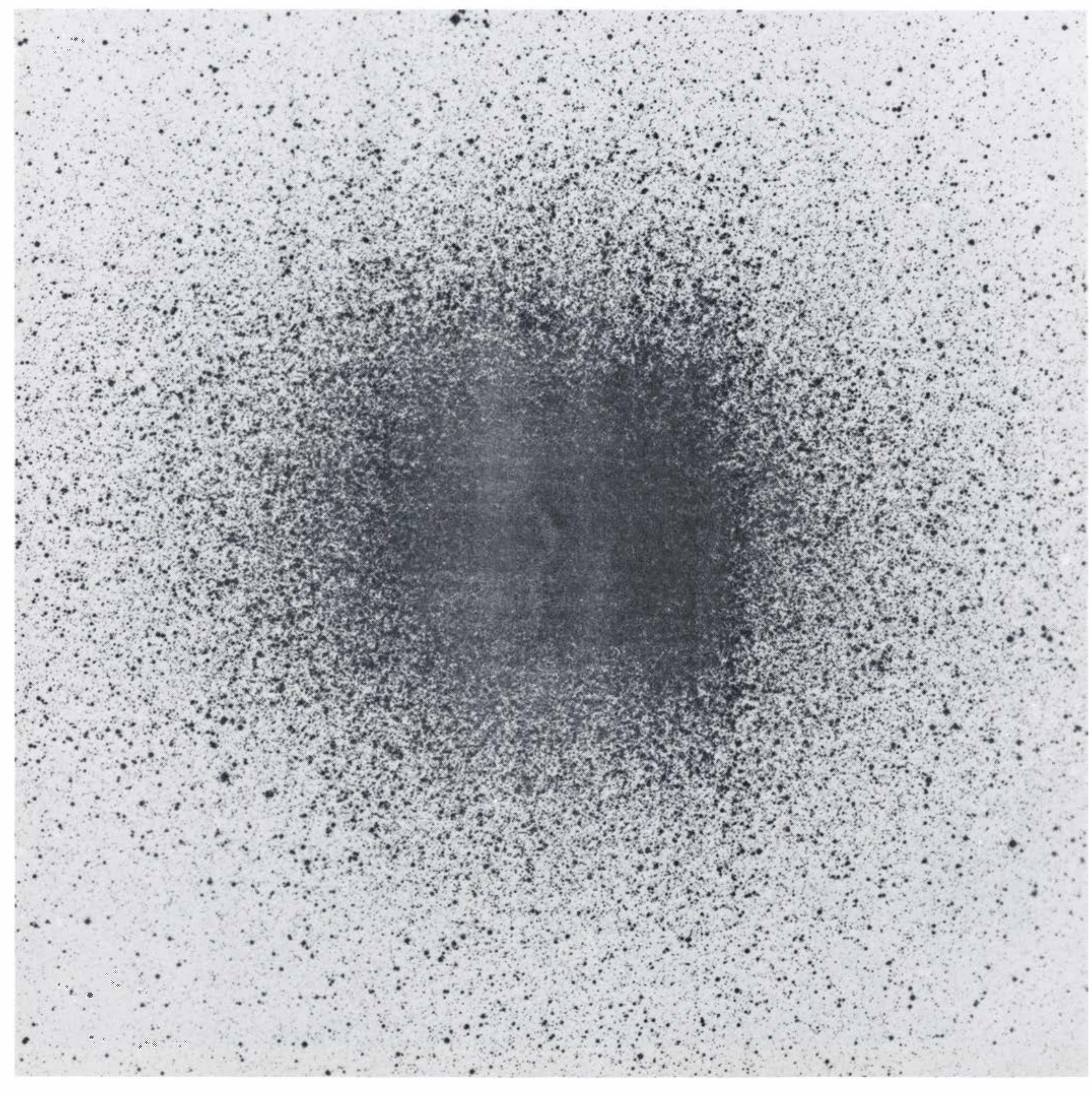
A alguien sorprenderá que el centro de la Vía Láctea sea al menos tan misterioso como la propia corona galáctica. La verdad es que se hace sólo ligeramente más visible para los observadores en la Tierra. Veinticinco años atrás, cuando las observaciones se podían realizar únicamente en longitudes de onda visibles y de radio, se sabía de la abundancia de tres clases de objetos en el núcleo galáctico. Todos eran antiguos. El primero, los cúmulos globulares; el segundo eran las variables RR de Lira. Son estrellas antiguas que alternativamente lucen y se debilitan con un período del orden de un día. El tercero son las nebulosas planetarias. La mayoría de ellas son estrellas antiguas, colapsadas, conocidas por enanas blancas rodeada, cada una, de una nube de gas que se cree es el despojo de la atmósfera de la estrella. El mismo centro galáctico estaba oculto a la visión óptica por capas y envolturas de polvo que



LA CONFIGURACION ESPIRAL MAS BRILLANTE en la vecindad del Sol es la Gran Nebulosa de Carina, que marca el lugar donde ocurre el encuentro entre el brazo de Sagitario y el brazo de Carina. La nebulosa es una nube de hidrógeno, ionizado (y por tanto luminoso) por la radiación ultravioleta de las estrellas supergigantes blancoazuladas recién formadas en el centro. Desde el Hemisferio Norte no se puede ver la nebulosa. La distancia que media hasta la nebulosa de Carina es de 2700 parsecs. La fotografía se realizó en el Observatorio del Monte Stromlo, adscrito a la Universidad Nacional Australiana.

se estima atenúan la luz del centro en cuantía tan grande como 30 magnitudes astronómicas, o en un factor de 10^{12}. El polvo era particularmente eficaz para filtrar la luz azul, a la cual presentaban su máxima sensibilidad las emulsiones fotográficas que se empleaban en los primeros períodos de la astronomía fotográfica. Parte del polvo estaba situado a lo largo de una senda en el margen del núcleo central, a unos pocos miles de parsecs del sistema solar. Y parte se encuentra a sólo unos 300 parsecs del sistema solar, en la constelación de Ofiuco, a lo largo de la visual desde la Tierra hacia el centro. Lo único que entonces se sabía acerca del centro es que albergaba una intensa radiofuente. La radiación electromagnética en la región de radio del espectro no la absorbía el polvo.

Hace unos 15 años, las perspectivas empezaron a mejorar. La detección de radioemisiones a una longitud de onda de 21 centímetros sirvió para delinear las fuentes de tal radiación: nubes de átomos neutros (no ionizados) de hidrógeno. La detección de emisiones a otras longitudes de onda reveló la presencia de nubes interestelares más oscuras, que constaban, principalmente, de moléculas. Sobre todo, el advenimiento de la astronomía de infrarrojo abrió una ventana al centro de la galaxia a longitudes de onda desde un micrometro hasta un milímetro. La radiación infrarroja constituye una segunda banda de longitudes de onda que el polvo interestelar no absorbe fuertemente. Una de las técnicas modernas más útiles es la detección de la radiación infrarroja emitida a una longitud de onda de 12,8 micrometros por los átomos de neón ionizados. Por ser un producto secundario de acontecimientos de alta



EL CUMULO GLOBULAR MAS BRILLANTE en la vecindad del Sol es Omega del Centauro, que se muestra aquí en un negativo de la fotografía tomada por Gary S. Da Costa con el telescopio de 2,5 metros del Observatorio de Las Campanas, ubicado en Chile. El cúmulo encierra cientos de miles de estrellas dentro de un diámetro de unos 30 parsecs. El cúmulo globular de Omega del Centauro dista del sistema solar algo más de 5000 parsecs.

energía, tales como las explosiones estelares, el neón se convierte en un constituyente ubicuo de las nubes interestelares; sirve, pues, de trazador.

Las observaciones a longitudes de onda de radio e infrarrojas producen cuatro clases de datos. En primer lugar, revelan la presencia de máximos locales en la lluvia de radiación procedente de ciertas direcciones del cielo. En segundo lugar, el corrimiento Doppler en las líneas espectrales de la radiación revela la velocidad radial de tal fuente, es decir, la componente de la velocidad de la fuente hacia el sistema solar o alejándose de él. En tercer lugar, el ensanchamiento de una banda espectral puede sugerir que la fuente está expandiéndose o contrayéndose. En cuarto lugar, las intensidades relativas de ciertas líneas en el espectro sugieren la temperatura de la fuente. Aun cuando toda la información posible haya sido extraída de la radiación, hay, sin embargo, muchas cosas que continúan ignorándose. Por ejemplo, no se puede averiguar si una fuente que está aproximadamente en la dirección del centro de la galaxia se encuentra delante o detrás de él, o si la velocidad radial de una fuente significa realmente que está girando alrededor del centro.

En su forma general, el núcleo central de la Vía Láctea es una esfera ligeramente achatada. Su borde exterior, que se encuentra a unos 5000 parsecs del centro, está marcado por un anillo de lo que ahora se llaman complejos moleculares gigantes. Son nubes interestelares grandes, oscuras y apelotonadas, constituidas principalmente por moléculas de hidrógeno. Volveré después sobre el tema. El núcleo mismo consta, en general, de una densa acumulación de estrellas antiguas inmersas en una matriz bastante delgada de gas interestelar y polvo. De la existencia de las estrellas se sabe por su radiación infrarroja, que se puede distinguir de la del gas o polvo. Una manera de explicar la relativa escasez de gas y polvo en el núcleo central es suponer que parte notable del mismo se condensó plácidamente hace mucho tiempo para formar las estrellas del núcleo. Por otro lado, varios aspectos interestelares del núcleo sugieren que el centro de la galaxia ha tenido una historia complicada y violenta.

De las características detectables del núcleo, la más externa es un anillo de hidrógeno neutro localizado a una distancia de 3000 parsecs del centro. El anillo fue descubierto en 1964 por Jan H. Oort y G. W. Rougoor, del Observatorio de Leiden. Los corrimientos Doppler de la radiación que emite muestran que está girando y, lo que es más importante, en expansión, con velocidades de alejamiento del centro que varían desde 50 hasta 135 kilómetros por segundo. Quizás el anillo sea un nuevo brazo espiral que se está desplegando. Pero resulta igualmente atractiva la hipótesis de que el centro de la galaxia expelió una especie de anillo de humo hace unos 30 millones de años. Es como si hubiera habido allí una explosión titánica. Tal vez la explosión barrió consigo gran parte del polvo y gas existente en el núcleo.

Prosiguiendo hacia su interior, hallamos otra característica distintiva a unos 1500 parsecs del centro. Para sus descubridores, Butler Burton, de la Universidad de Minnesota, y Harvey S. Liszt, del Observatorio Nacional de Radioastronomía, se trata de un disco constituido por hidrógeno, tanto en estado atómico como molecular. También está en rotación al mismo tiempo que se expande. Sorprendentemente, la mejor manera de interpretar los datos es suponiendo que el disco está inclinado en un ángulo de unos 15 a 20 grados con respecto al plano de la galaxia.

Composición del núcleo

Alguien podría esperar que la composición del núcleo central fuese más o menos homogénea desde el anillo de Burton-Liszt hacia el centro, aunque no fuese más que para simplificar la tarea con la que se enfrentan quienes tratan de comprender la estructura del núcleo. Pero nos aguardan más sorpresas. Otro anillo de humo existe, evidentemente, a unos 300 parsecs del centro; también es una mezcla de complejos moleculares, nubes de polvo y regiones de hidrógeno atómico y molecular. El hidrógeno atómico se encuentra ionizado en algunos lugares, lo que significa que está muy caliente: bastante por encima de los 10.000 grados Kelvin. Asociados a esas zonas calientes, hallamos cúmulos de estrellas supergigantes blancoazuladas y de origen reciente. ¿Por qué tienen que estar esos reinos de alta temperatura y formación de estrellas precisamente en este anillo? Lo que le deja a uno especialmente perplejo es la existencia de otro anillo más frío de átomos sólo ligeramente ionizados a una temperatura de 5000 grados y que se encuentran a no más de 10 parsecs del centro. A 10 parsecs, el anillo aparece bastante denso y en rotación.

Los tres parsecs centrales de la galaxia incluyen, evidentemente, varios millones de estrellas; dan al centro la más densa acumulación de estrellas de toda la galaxia. La región nuclear encierra también cierto número de nubes compactas de gas ionizado; un grupo de investigadores dirigido por John H. Lacy y Charles H. Townes, de la Universidad de California en Berkeley, ha detectado 14 de ellas. Una nube típica tiene aproximadamente la masa del Sol y un diámetro de una fracción de parsec, y está girando rápidamente alrededor del centro; completa su órbita en unos 10.000 años (compárese con el período orbital del Sol, de unos 200 millones de años). Luis Rodríguez y Eric J. Chaisson, del Observatorio del Harvard College, han puesto de manifiesto que la velocidad del gas ionizado aumenta con la proximidad al centro. Todo esto sugiere que las nubes son satélites de un objeto interior supermasivo.

Cualquiera que sea lo que haya en el propio centro, aparece como una brillante fuente infrarroja en los mapas hechos por Eric E. Becklin y Gerry Neugebauer, del Instituto de Tecnología de California. Según Bruce Balick, de la Universidad de Washington, y Robert L. Brown, del Observatorio Nacional de Radioastronomía, que han estudiado las radioemisiones del objeto central, tiene un diámetro no superior a 10 veces la distancia de la Tierra al Sol. La dimensión de su masa puede llegar a 50 millones de veces la solar. La conjetura que estimamos más probable es aquella que sostiene que el propio centro de la Vía Láctea aloja un agujero negro creado por la caída hacia dentro de cientos de miles de estrellas. El centro vendría a ser un cementerio estelar.

Estructura espiral óptica

Considerados la periferia y el centro de la Vía Láctea, resta por abordar una región intermedia que tiene en su parte central el sistema solar. Se trata de la fracción de la galaxia donde predomina la estructura espiral.

El rastreo científico de los brazos espirales de la Vía Láctea empezó hace unos treinta años, cuando Morgan, Osterbrock y Sharpless distinguieron tres segmentos de brazos espirales. Para reconstruirlos, representaron las posiciones de las estrellas supergigantes blancoazuladas, clasificadas a partir de su configuración de líneas espectrales como estrellas *O* y *B*, junto con las brillantes nubes de átomos de hidrógeno ionizado que frecuentemente rodean a tales estrellas. Sus diagramas muestran un brazo de Orión, del que el Sol es un miembro; un brazo de Perseo, a 2000 parsecs más allá del centro de la galaxia, y un brazo de Sagitario, 2000 par-

secs más próximo al centro de la galaxia. El nombre de los brazos refleja la práctica corriente en astronomía de emplear las constelaciones para indicar direcciones en el cielo. A los tres brazos identificados por Morgan y sus colegas se ha añadido, desde entonces, el brazo de Carina, que puede ser una continuación del brazo de Sagitario; la concatenación se llama brazo de Sagitario-Carina. Los dos segmentos que lo componen se encuentran en la Gran Nebulosa de Carina, que comprende un gran número de estrellas *O* y *B*. Existen también indicios de otros segmentos.

Un trabajo reciente, realizado por Roberta M. Humphreys, de la Universidad de Minnesota, confirma que las estrellas *O* y *B* abundan en los brazos principales, reconocidos hace treinta años. Una buena noticia, sin duda. Significa que los brazos espirales están realmente delineados por estrellas supergigantes blancoazuladas y muy calientes, por los cúmulos formados por tales estrellas y por las brillantes nubes de gas donde se encuentran tales cúmulos y estrellas. Las estrellas *O* y *B* son muy jóvenes; las que se ven ahora en los segmentos de brazos espirales de la Vía Láctea se formaron no hace más de 10 millones de años. No hay duda de que los brazos espirales son regiones de formación de estrellas.

Los márgenes interiores de la mayoría de los brazos espirales están delimitados, según parece, por nebulosas oscuras, que son nubes frías de átomos, moléculas y polvo. Según la teoría de la onda de densidad de Lin y Shu, las nebulosas oscuras señalan la compresión de la materia interestelar por una onda de presión que avanza a través del disco galáctico. La compresión precipita la condensación de materia para formar las estrellas del segmento espiral. Podría esperarse que las estrellas jóvenes estuvieran situadas detrás de la onda de densidad que avanza y que las nebulosas oscuras estuvieran dentro de ella. Sin embargo, de acuerdo con la teoría, la onda se mueve solamente a dos tercios de la velocidad de rotación del disco galáctico. De aquí que la onda sea adelantada por las estrellas que se han formado en su interior, y que la posición de las nebulosas oscuras venga a situarse en el borde posterior de las estrellas.

Algunos de los resultados de Humphreys, todo hay que decirlo, son menos alentadores. Entre las estrellas llamadas Cefeidas variables, las hay tan jóvenes como las estrellas *O* y *B;* en particular, las Cefeidas variables que alternativamente se iluminan y se debilitan, con un período superior a 15 días, tienen edades de no más de 10 millones de años. Además, son también grandes y brillantes, de modo que pueden verse a largas distancias. Por ser jóvenes, no deben haberse alejado mucho de su lugar de nacimiento, presumiblemente una nebulosa oscura en el borde interior de un brazo espiral. Teniendo en cuenta todo ello, las Cefeidas variables de largo período deberían ser excelentes señalizadoras de la estructura espiral. Pero no lo son. Las Cefeidas variables representadas por Humphreys forman una distribución de puntos en el plano galáctico esencialmente caótica.

Hay una cosa que siempre deben tener presente quienes tratan de descubrir la estructura espiral en las observaciones ópticas de la Vía Láctea. El método corriente de calcular la distancia de una estrella consiste en comparar su luminosidad observada con lo que se supone es su luminosidad intrínseca. Esta se asigna a partir de la distribución de líneas en el espectro de la radiación de la estrella. La disminución del brillo guarda correlación con la distancia. Pero, incluso empleando las mejores técnicas, se obtiene la distancia de una estrella del tipo *B* u *O* con una incertidumbre en más o en menos del 10 por cien. Por ejemplo, una estrella que se calcula estar a 2700 parsecs del Sol –digamos una estrella en la nebulosa de Carina– puede hallarse en realidad a sólo 2400 parsecs o 3000 parsecs. El resultado es un emborronamiento puramente observacional de varios cientos de parsecs en la representación de lo que puede constituir una estructura espiral bien definida.

Con los más modernos telescopios, espectrógrafos y equipo fotográfico, se pueden detectar estrellas *O* y *B* y examinar sus espectros a grandes distancias, que se calculan de hasta 8000 parsecs del Sol. A esa distancia, sin embar-

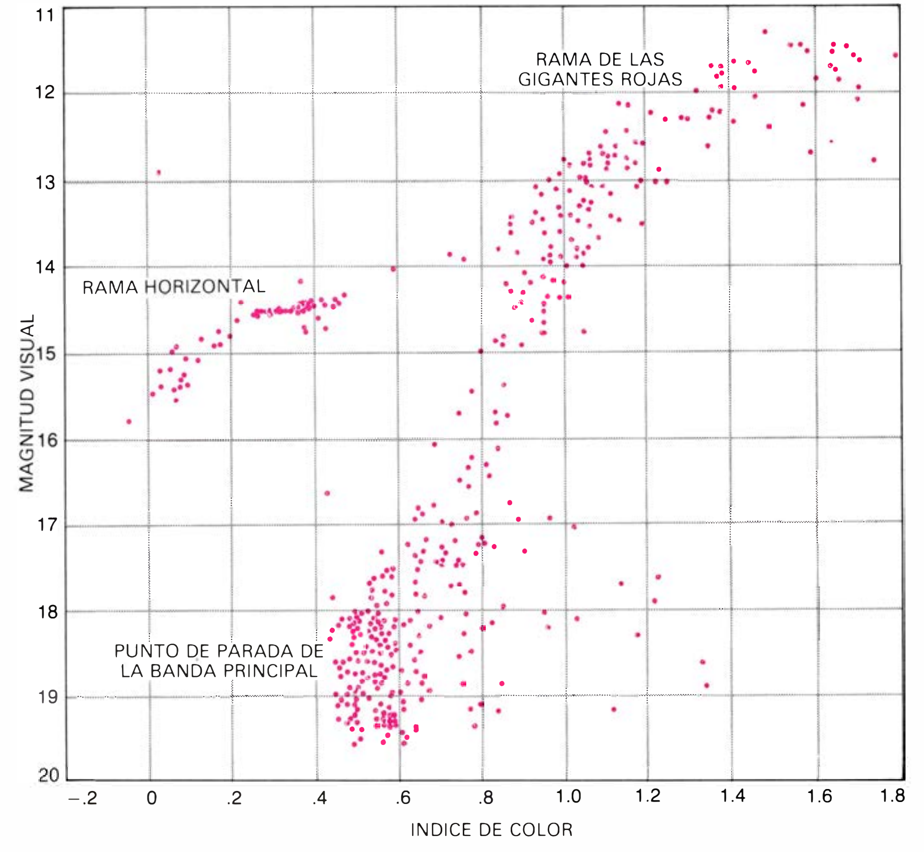


DIAGRAMA DE HERTZSPRUNG-RUSSELL del cúmulo globular Omega del Centauro, que nos sugiere una curiosa composición para las estrellas que lo forman. La rama del diagrama constituida por estrellas gigantes rojas (*parte superior derecha*) es más larga de lo usual, lo que significa un insólito grado de variación en el contenido de "metales" (átomos más pesados que el carbono) de las gigantes rojas. En concreto, los átomos de metal de una estrella aumentan su opacidad a la radiación que escapa del interior. Consiguientemente, cambian su color: las gigantes rojas ricas en metales son más rojas y, en un diagrama de Hertzsprung-Russell, su posición queda sesgada hacia la derecha. La variación en el contenido del metal sugiere, a su vez, que las estrellas del cúmulo de Omega del Centauro se formaron a lo largo de más de un único episodio. Las deducciones ulteriores son problemáticas. En general, la edad de un cúmulo globular se determina así: comparando la luminosidad y el color de las estrellas en lo que se llama parada de la banda principal en el diagrama de Hertzsprung-Russell (*abajo*) con las de otras estrellas, sitas en una posición correspondiente, en diagramas generados en simulaciones por ordenador, que recrean la evolución del cúmulo. En el caso de Omega del Centauro, los datos no revelan la extensión total de la banda principal, de modo que es difícil situar con precisión el punto de parada. Esto hace difícil establecer la edad o edades del cúmulo. Los datos fueron recogidos por Cannon y N. J. Stewart, de Edimburgo.



CURIOSO NACIMIENTO DE ESTRELLAS, visible en la pequeña nube oscura, a la izquierda de la fotografía. La nube pertenece a la clase llamada glóbulos. Mide medio parsec de diámetro y, probablemente, su masa no llega a 100 veces la solar. La masa del glóbulo está compuesta de moléculas (principalmente hidrógeno) y polvo. En este glóbulo particular, hay un par de estrellas incipientes y unidas por un filamento luminoso que han sido evidentemente expulsadas de él. El filamento cruza el borde superior de la nube. La expulsión pone en entredicho la hipótesis según la cual un glóbulo se condensa en aproximadamente un millón de años para formar una estrella nueva y única. La línea brillante que se extiende diagonalmente desde la parte inferior hacia la derecha del campo es el borde de la nebulosa Gum, que se cree que es, en parte, restos de una explosión de supernova. Se calcula que tanto el glóbulo como la nebulosa están aproximadamente a 300 parsecs del sistema solar. Se encuentran en la parte del cielo meridional marcada por la constelación Vela. La fotografía fue tomada por el autor en 1978 con el telescopio de cuatro metros del Observatorio Interamericano de Cerro Tololo, en Chile.

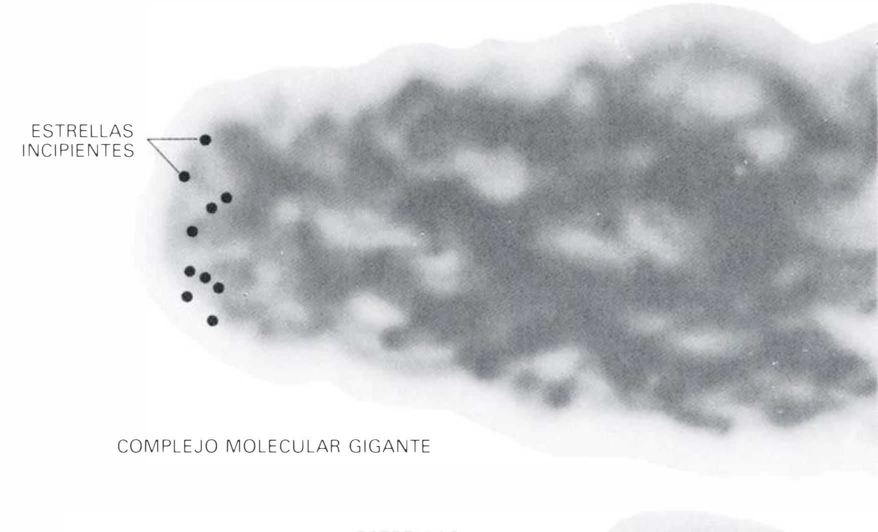
go, la inexactitud en el cálculo es de más o menos 800 parsecs. En tal caso, una estructura espiral puede muy bien resultar irreconocible. Parece haber poca esperanza de reconstruir la estructura espiral de la Vía Láctea por medios ópticos a distancias superiores a los 8000 parsecs.

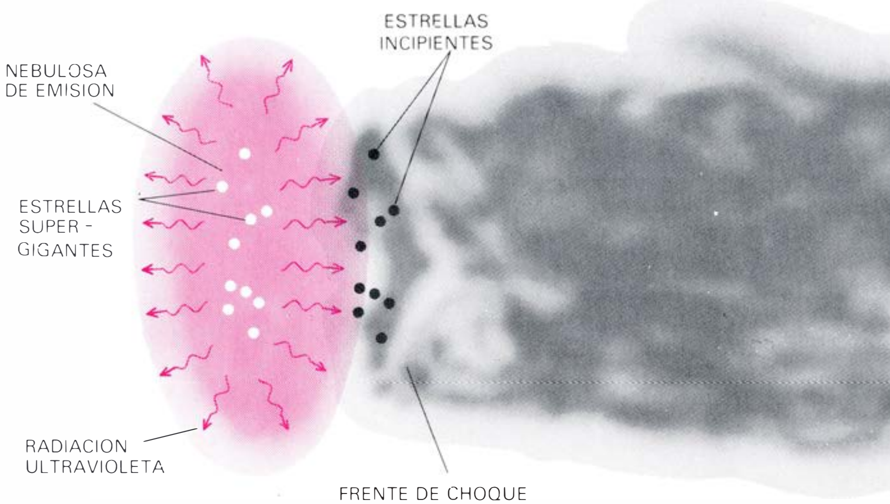
Radioestructura espiral

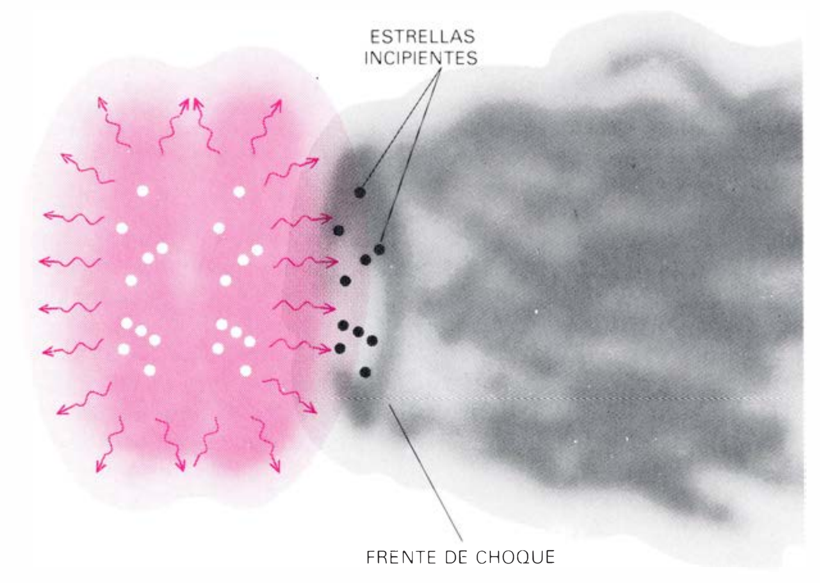
¿Qué frutos han cosechado los esfuerzos encaminados a reconstruir la estructura espiral cuando se emplean longitudes de onda de radio? De nuevo, las perspectivas están lejos de ser alentadoras. A comienzos del decenio de 1950, cuando Morgan, Osterbrock y Sharpless presentaron las características ópticas clave de la estructura espiral, Harold I. Ewen y Edward M. Purcell, de la Universidad de Harvard, habían detectado radiación procedente de nubes interestelares de hidrógeno atómico neutro, a una longitud de onda de 21 centímetros. En el curso de unos pocos años, estaban disponibles los primeros radiomapas de la galaxia. Estos sugerían que, al menos en la parte exterior del disco galáctico, las nubes de hidrógeno atómico neutro estaban dispuestas en una configuración que semejaba una espiral circular.

Vale la pena examinar algunos detalles de la manera como se acumulan e interpretan tales datos. Cuando un radiotelescopio detecta radiación en, y cerca de, una longitud de onda de 21 centímetros, procedente de una pequeña región del cielo, está en realidad recibiendo radiación proviniente de cierto número de nubes de hidrógeno atómico neutro, situadas a lo largo de una visual única. Cada nube posee su propia velocidad de acercamiento o alejamiento con respecto al telescopio, y de aquí que su radiación presente un corrimiento Doppler característico. Como resultado, el telescopio apuntado en una dirección única recibe un perfil de crestas y valles cuando se representa en un gráfico de intensidad frente a longitud de onda. La onda indicadora de 21 centímetros es, en realidad, un conjunto de líneas muy juntas con diferentes intensidades, diversos grados de ensanchamiento y distintos corrimientos Doppler. Como se supone que el hidrógeno atómico neutro presenta su mayor densidad en el borde posterior de los brazos espirales, parece razonable pensar que debería haber un máximo de gran intensidad en el perfil para la longitud de onda que corresponde a la velocidad de acercamiento o alejamiento del hidrógeno donde la visual corta un brazo espiral. Con la ayuda de una curva de rotación de la Vía Láctea, se podría entonces determinar la distancia a dicho accidente.

Esta manera de razonar, sencilla y atractiva, no resiste un análisis serio.







SE CREE QUE EL NACIMIENTO DE ESTRELLAS SUPERGIGANTES acontece en fases repetitivas. La secuencia comienza (*parte superior*) con la formación de un cúmulo de estrellas supergigantes cerca de la periferia de un complejo molecular gigante, región fría y oscura de moléculas (principalmente hidrógeno) y polvo. Después de algunos millones de años, la radiación ultravioleta de las nuevas estrellas ha ionizado el hidrógeno cercano, de modo que las estrellas están envueltas en una nebulosa de emisión o nube de gas luminoso. Además, la presión de radiación ha comprimido la materia en el complejo molecular gigante. (Cada cuanto de radiación posee cantidad de movimiento; empuja, pues, las moléculas y el polvo.) De ello resulta la formación de un segundo grupo de estrellas supergigantes (*dibujo del centro*). Transcurridos otros cuantos millones de años, una sucesión similar puede originar un nuevo grupo de estrellas (*dibujo inferior*). El diagrama se basa en una hipótesis propuesta por Bruce G. Elmegreen, de Columbia, y Charles J. Lada, de Arizona, quienes han dado la más clara explicación de lo que sucede en estos casos. La hipótesis justifica la presencia de cúmulos de estrellas supergigantes en nebulosas, tales como la Gran Nebulosa de Orión, que están situadas en la vecindad de un complejo molecular.

La primera complicación sobreviene de las nubes de hidrógeno atómico neutro, que tienen movimientos propios, autónomos completamente respecto del movimiento de las estructuras espirales. Esos movimientos independientes pueden fácilmente alterar la velocidad total de una nube en una cuantía de hasta seis kilómetros por segundo. Otra complicación procede de la materia del disco galáctico: al parecer, presenta flameos en gran escala. Burton y sus colaboradores, del Observatorio Nacional de Radioastronomía, han demostrado que bastan ligeros movimientos de esta clase para provocar máximos de intensidad en la radiación de 21 centímetros para direcciones en las que la visual no corta ninguna configuración espiral. Todavía más, hay direcciones en el cielo para las que la velocidad de acercamiento o alejamiento de una nube de gas respecto al sistema solar puede cambiar sólo muy lentamente con la distancia. El perfil de 21 centímetros puede entonces mostrar un máximo causado por intervención de la anchura de una única gran extensión de gas de densidad uniforme.

Se pueden ensayar otros varios enfoques para abordar la estructura espiral. Por ejemplo, observar las líneas espectrales emitidas por moléculas de monóxido de carbono a longitudes de onda próximas a 2,6 milímetros. Las líneas de monóxido de carbono dan acceso a una clase de nubes que son más frías que las compuestas de hidrógeno atómico neutro. Las nubes más frías están constituidas, principalmente, por hidrógeno molecular, con alguna adición de monóxido de carbono y otras sustancias. (El principio general es que el aumento de temperatura conduce primero a la disociación de moléculas en átomos neutros y después a la ionización de los átomos.) William Herbst, de la Wesleyan University ha demostrado que la presencia de monóxido de carbono en una nube interestelar está correlacionada con la presencia de polvo. Las líneas de monóxido de carbono ayudan, por tanto, en la detección de distantes nubes de polvo y complejos de polvo que de otra manera podrían escaparse a la observación.

Una exploración a baja resolución de las nubes de monóxido de carbono, realizada por Richard S. Cohen y Patrick Thaddeus, del Instituto Goddar para Estudios Espaciales, y Thomas M. Dame, de la Universidad de Columbia, aporta señales de la existencia de estructuras en espiral. Así, en algunos casos las nubes de monóxido de carbono parecen llenar huecos entre nubes de

hidrógeno atómico en algunas de las estructuras espirales reconocidas. Sin embargo, dudo que se hubiera descubierto la estructura espiral de la Vía Láctea si los únicos datos disponibles hubieran sido las radioobservaciones a longitud de onda de 2,6 milímetros.

Me desagrada profundamente tener que mostrarme pesimista ante una posible cartografía de la estructura espiral de la Vía Láctea más allá de una distancia de unos 8000 parsecs del Sol, sobre todo si advertimos que hasta esta distancia la estructura espiral parece firmemente establecida; no veo cómo podría mejorarse en los próximos diez o veinte años. No quiero decir con ello que el estudio de la estructura espiral se encuentre en punto muerto. Entre los telescopios ópticos, los grandes reflectores ya en uso y el Telescopio Espacial, instrumento que se colocará en órbita alrededor de la Tierra, prometen revelar los pequeños detalles de la estructura espiral y los movimientos de características en pequeña escala. Entre los radiotelescopios, el Sistema Westerbork en Holanda y la Gran Distribución en Nuevo México, que sintetizan imágenes a partir de un conjunto de detectores, harán progresar la investigación de las frías y oscuras nubes de las características (o accidentes) espirales y sin duda sugerirán muchas ideas acerca de las causas de la estructura espiral. Tales prometedores progresos se reconocen, no obstante y de manera principal, en galaxias espirales distintas de la nuestra. Las galaxias espirales Messier 31, 33, 51, 81 y 101 constituyen algunos ejemplos relativamente cercanos.

Las edades de los cúmulos globulares

Un campo de investigación que está floreciendo en el dominio de la astronomía galáctica y en el de la extragaláctica es el examen de los cúmulos globulares. Parte de la importancia de los cúmulos reside en que, al parecer, se trata de los objetos más antiguos de la Vía Láctea. Apuntan, por tanto, al nacimiento y evolución de la galaxia y, por ende, al del primitivo universo. La hipótesis más sencilla acerca de los cúmulos es la que sostiene que todos se formaron dentro de un corto período de tiempo (unos mil millones de años) después de la gran explosión, o instante en que toda la materia del presente universo se formó explosivamente, según se cree, partiendo de un punto único. Los cúmulos se contarían entonces entre los primeros objetos que se condensaron a medida que las galaxias tomaban forma, evolucionando, cada uno, a partir de una burbuja de gas. A favor de esta hipótesis reza el que la mitad de los aproximadamente 200 cúmulos globulares de la Vía Láctea estén distribuidos por todas partes en el volumen casi esférico del halo galáctico. Presumiblemente, nacieron allí mucho antes de que el disco galáctico tomara forma. La órbita de un tal cúmulo es típicamente una elipse bastante excéntrica, en tanto que las órbitas circulariformes son características de la materia del disco. El eje mayor de la órbita mide varias decenas de miles de parsecs; una vez cada mil millones de años, la órbita del cúmulo cruza el espesor del disco, que mide sólo unos cuantos cientos de parsecs.

En la hipótesis más sencilla, las estrellas de los cúmulos globulares se habrían formado en una época en que la materia disponible en la galaxia era principalmente hidrógeno y helio, los dos elementos químicos que se suponen haber sido creados inmediatamente después de la gran explosión. En contraste, una estrella nacida más tarde se habría condensado a partir de gas interestelar, una fracción del cual habría pasado por el interior de las estrellas de origen precedente. Parte del gas, por ejemplo, habría sido expulsado al espacio por explosiones de supernova. Se trataría de materia en la que se habrían creado átomos más pesados por fusión termonuclear. La estrella de formación tardía tendría, por tanto, mayores concentraciones de elementos químicos más pesados; en el lenguaje abreviado de los astrofísicos, todos estos elementos se conocen como metales.

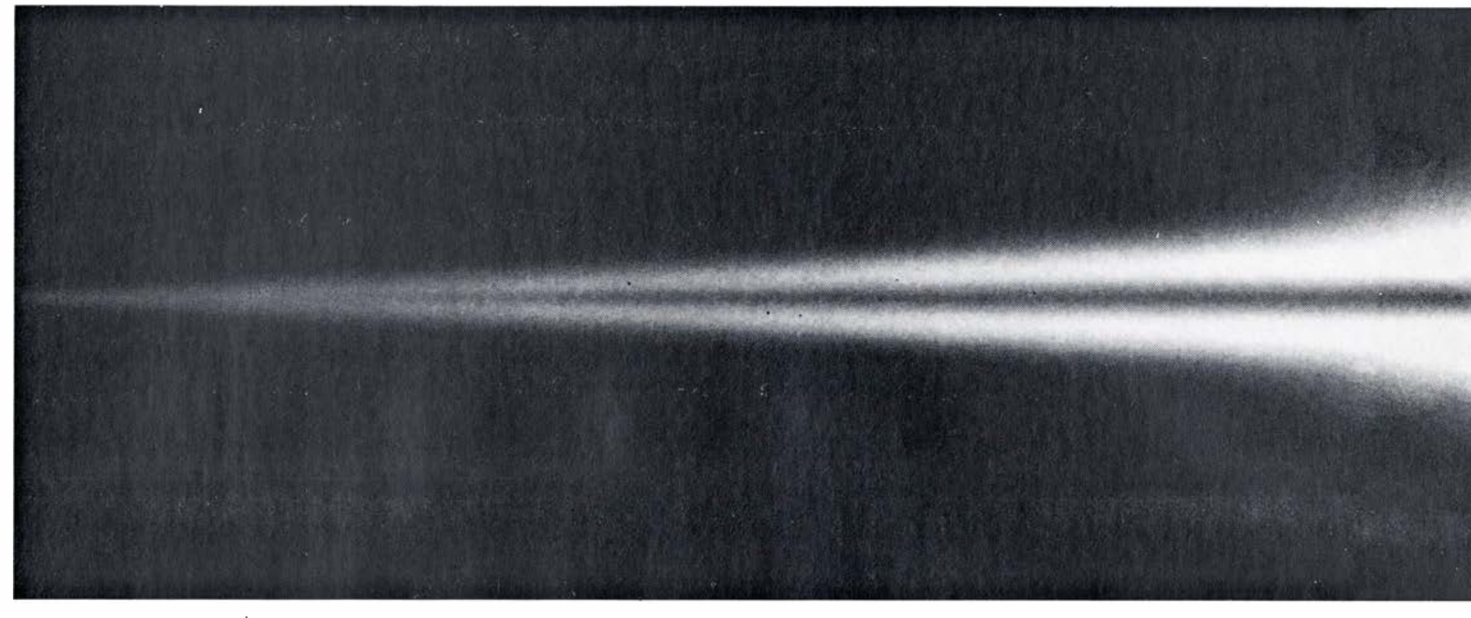
Pero es, justamente, la concentración de metales en los diversos cúmulos globulares lo que pone en peligro esta sencilla hipótesis. La relación de metales a hidrógeno y helio es 100 veces mayor en el Sol que en las estrellas de los cúmulos globulares pobres en metales, tales como M3 (de Messier 3). Además, estos cúmulos pobres en metales tienden a identificarse con los que están situados en la parte más externa de la galaxia. Se les asigna ahora una edad de unos 15.000 millones de años. Por otro lado, el cúmulo globular 47 Tucanae es relativamente rico en metales. Su edad, se cree, es de 10.000 millones de años, el doble de la edad del Sol. Omega del Centauro, el cúmulo globular más impresionante de la Vía Láctea, muestra múltiples concentraciones de metales. Se trata de una aglomeración idiosincrásica de estrellas que nacieron en diferentes épocas. La dispersión de edades de los cúmulos globulares está en contradicción con los modelos actuales acerca de la evolución de la galaxia. Ningún modelo le concede a la condensación del disco galáctico un período de tiempo cuya magnitud se cifre en varios miles de millones de años.

Con respecto a la cuestión de la edad, debe notarse que la edad máxima del universo puede deducirse a partir de las velocidades a las que las galaxias se están separando unas de otras. El valor actual para el máximo se acerca a los 15.000 millones de años. Investigaciones recientes de la velocidad de separación se muestran proclives a rebajar la edad máxima. Si esta tendencia continúa, la edad podría reducirse verosímilmente hasta sólo 10.000 millones de años. Habría que explicar entonces por qué ciertos cúmulos globulares parecen exceder la edad asignada al universo. Importa, pues, que se calculen valores fidedignos para las edades de los cúmulos globulares. Y convendría que tales edades estuvieran disponibles a finales de esta década, porque, para entonces, el Telescopio Espacial deberá haber proporcionado valores, mucho más dignos de confianza, para las velocidades de separación y las distancias de las galaxias. De esta manera, habrá suministrado un valor mucho más ajustado para la edad máxima del universo.

Evolución de los cúmulos globulares

La dinámica interna de los cúmulos globulares queda bien representada por tres tiempos característicos. El primero es el tiempo de cruce, o tiempo que tarda una estrella en moverse, a través del cúmulo, bajo la atracción gravitatoria del cúmulo como un todo. El segundo es el tiempo de relajación, o tiempo que emplea la estrella en asentarse y convertirse en un miembro estable del cúmulo bajo la influencia de sus interacciones gravitatorias con las estrellas vecinas. El tercero es el tiempo de evolución, o tiempo en el que una estrella estable cambia su forma y composición de manera significativa. Para un cúmulo globular rico en estrellas, el tiempo de cruce es mucho más corto que el de relajación, que, a su vez, es mucho más corto que el tiempo de evolución.

El trabajo de Ivan R. King, de la Universidad de California en Berkeley, ha sugerido algunas puntualizaciones más. Cuando un cúmulo globular rico en estrellas lleve existiendo unos cin-



VISTA EXTERIOR DE LA VIA LACTEA, donde aparece el borde de la galaxia como la vería un astrónomo situado en otra galaxia y provisto de un telescopio. La imagen fue generada con la ayuda de un ordenador por John N. Bahcall y Raymond M. Soneira, del Instituto de Estudios Superiores. El

cuenta tiempos de cruce, sus estrellas se habrán asentado; la estadística de sus velocidades se parecerá mucho a las de las velocidades de las moléculas de una nube de gas. Estado de equilibrio que podría persistir indefinidamente si las estrellas nunca cambiasen internamente y si ninguna de ellas abandonara el cúmulo.

Pero ya por los años 30, Lyman Spitzer, Jr., de Princeton, y Victor A. Ambartsumian, del Observatorio Astrofísico Byurakan, en Armenia, habían llegado a la conclusión de que las estrellas de menor masa en un cúmulo eran probablemente las que tenían una velocidad mayor y que, en muchos casos, las estrellas de máxima velocidad escaparían del cúmulo. A causa de su gran velocidad, las estrellas que escapasen se llevarían consigo una porción de la energía total del cúmulo mayor que el promedio por estrella. Como resultado, el cúmulo se contraería. A lo largo de un intervalo del orden de lo que hemos definido tiempo de evolución, el cúmulo perdería una parte apreciable de su energía media por estrella, y las estrellas más masivas se irían aproximando al centro.

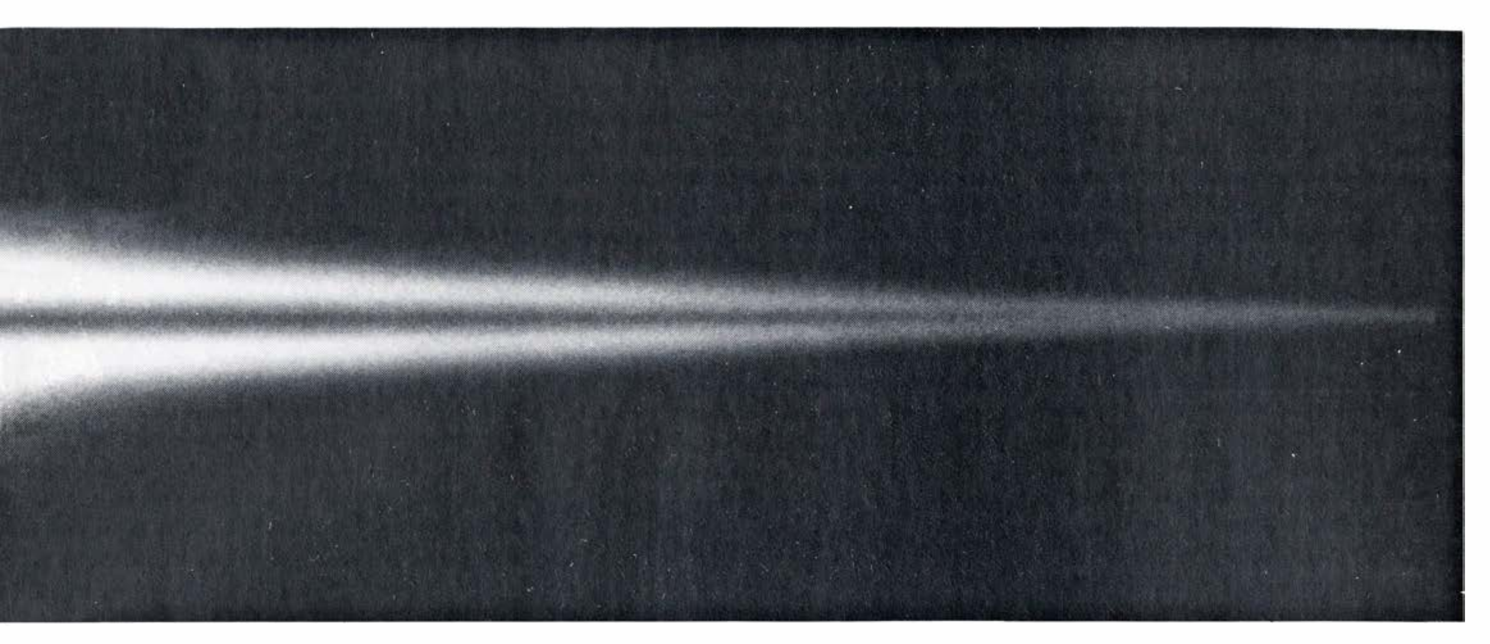
Los cambios internos entre las estrellas que quedan sólo pueden acelerar esta tendencia. Digamos de entrada que la pérdida de masa por las estrellas constituye un hecho corriente en astrofísica, según ha podido averiguarse. Al comienzo de su evolución, una estrella típica está inmersa en una envoltura de gas, que se está gradualmente expeliendo por la presión, dirigida hacia fuera, de la radiación electromagnética de la estrella. (Cada fotón, o cuanto de radiación, tiene cantidad de movimiento, y por tanto empuja a lo que le absorbe.) Más tarde, la estrella comienza a expulsar un débil viento estelar de partículas. En un momento ulterior, ciertas clases de estrellas explotan. En una nova, la estrella que explota lanza al espacio el equivalente a la masa de un planeta del tamaño de Júpiter. Se llama supernova a la explosión en la cual se destruye la totalidad de una estrella. Al producirse pérdida de masa mediante mecanismos como los descritos, las estrellas de un cúmulo globular se van corriendo sistemáticamente hacia clases de masa menor (y brillo intrínseco más débil). A lo largo de un intervalo del orden del tiempo de relajación, adquieren la mayor velocidad de una estrella de menor masa. De aquí que aumente su probabilidad de escapar del cúmulo.

En los miles de millones de años que un cúmulo puede pasar en el delgado halo exterior (o incluso en la corona) de la Vía Láctea, la evolución del cúmulo queda afectada tan sólo por los procesos internos que acabo de describir. Llega un momento, sin embargo, en que el cúmulo atraviesa el disco galáctico. En ese instante, el cúmulo corre el riesgo, no tanto de choque, cuanto de interacción gravitatoria con la materia situada en y cerca del plano galáctico. La fuerza de la interacción puede desgarrar el cúmulo.

A medida que la órbita del cúmulo conduce a éste más cerca del centro de la galaxia, se va perfilando una nueva amenaza. En efecto, la masa central de la galaxia ejerce una fuerza atractiva sobre el lado del cúmulo que pasa más cerca del centro mayor que sobre el otro lado. La diferencia entre la fuerza más interna y la más externa puede deformar el cúmulo. De hecho, existe un radio crítico para cualquier cúmulo globular: la distancia del centro galáctico dentro de la cual algunas estrellas del cúmulo serán arrancadas de éste cuando pase a la mínima distancia del centro. La distancia crítica se llama radio de marea, porque el mecanismo que deforma el cúmulo es semejante al que eleva las mareas en los océanos. No ha de sorprendernos que los cúmulos globulares del interior del núcleo central de la galaxia sean, por término medio, menores que los que están en el halo o en la corona galáctica.

Un último factor puede afectar a los cúmulos globulares. Lo menciono con cierto recelo. No hará más de 15 años, pocos astronómos hubieran creído que la formación de sistemas de estrellas dobles fuera importante en los cúmulos de estrellas. Por ese entonces, Sebastian von Hoerner, del Observatorio Nacional de Radioastronomía, llevó a cabo la primera simulación mediante ordenador de cúmulos abiertos, o agregaciones de estrellas que difieren de los cúmulos globulares en que las estrellas son jóvenes, en menor número y más dispersas. Von Hoerner basó la simulación en ecuaciones que representaban las interacciones gravitatorias entre grupúsculos de estrellas. Las simulaciones mostraron una y otra vez que las estrellas de un cúmulo tendían a unirse en parejas para originar sistemas binarios.

En simulaciones de ordenador más modernas, que incorporaban primero 100, después 500 y hasta 1000 estrellas en interacción mutua, se notó el mismo efecto. En 1975, Douglas C. Heggie, de la Universidad de Edimburgo, mostró



comienzo de la simulación fue una descripción matemática del número de estrellas de brillo dado en una dirección determinada, a partir del sistema solar. El ordenador representó la configuración resultante de luminosidad (y una vía central de polvo), tal como la vería un observador situado fuera.

que un sistema binario recién formado constituía un formidable sumidero de energía disponible en el cúmulo. A grandes rasgos, la formación de cada sistema binario extrae energía cinética del movimiento de las estrellas y la convierte en energía que liga las dos estrellas en órbitas, cada una alrededor de la otra. Consiguientemente, la pareja se va hundiendo hacia el centro del cúmulo. Todavía más recientemente, Spitzer y sus colaboradores en Princeton han desarrollado modelos que sugieren que en la densamente poblada región central de un cúmulo globular, donde se arraciman las estrellas de mayor masa, cada estrella capturará, probablemente, una o incluso dos o más compañeras. Como resultado, la región central puede llegar a estar todavía más densamente apretada a medida que el cúmulo globular evoluciona.

La razón de mi recelo a la vista de todas estas pruebas circunstanciales es sencilla. Los sistemas binarios masivos de estrellas predominan en los cúmulos abiertos. Tenemos en Mizar un ejemplo clásico. Se trata de una estrella en el cúmulo abierto de la Osa Mayor; el observador la reconocerá mejor como la estrella que aparece en el codo del asta del Carro. Mizar tiene una compañera llamada Alcor, que se deja contemplar por quienes gozan de excelente vista. Mizar, se concluirá, es un sistema binario. Pero la verdad es que tanto Mizar como Alcor tienen estrellas compañeras. Cada una de ellas es un sistema binario. Lo que convierte en todavía más exasperante la búsqueda de ejemplos de sistemas masivos de estrellas binarias en la vecindad del centro de los cúmulos globulares, es que no ha tenido siempre el mismo éxito. En realidad, no ha tenido éxito alguno.

Pero sí se ha encontrado algo distinto. En 1976, Jonathan E. Grindlay y Herbert Gursky, del Centro de Astrofísica del Observatorio del Harvard College y del Observatorio Astrofísico Smithsoniano, informaron que una explosión de rayos X de terrorífica intensidad había alcanzado el sistema solar procedente de una dirección cercana a la del centro del cúmulo globular NGC 6624. El brote duró solamente 8 a 10 segundos. Desde entonces, se han detectado repetidamente otros brotes semejantes procedentes de seis cúmulos globulares. Según una de las hipótesis, existe un agujero negro con una masa no menor de 100 masas solares en el centro de cada uno de tales cúmulos. Los brotes o explosiones resultarían entonces del choque intermitente del gas interestelar con un disco de deposición gaseoso y caliente que rodea al agujero negro. Viene enseguida a la mente que el hipotético agujero negro podría ser la consecuencia última del colapso gradual de las estrellas hacia el centro del cúmulo, colapso al cual la formación de sistemas binarios puede haber aportado su contribución y quizás haberle dado los empujones finales.

Nacimiento de estrellas en complejos moleculares

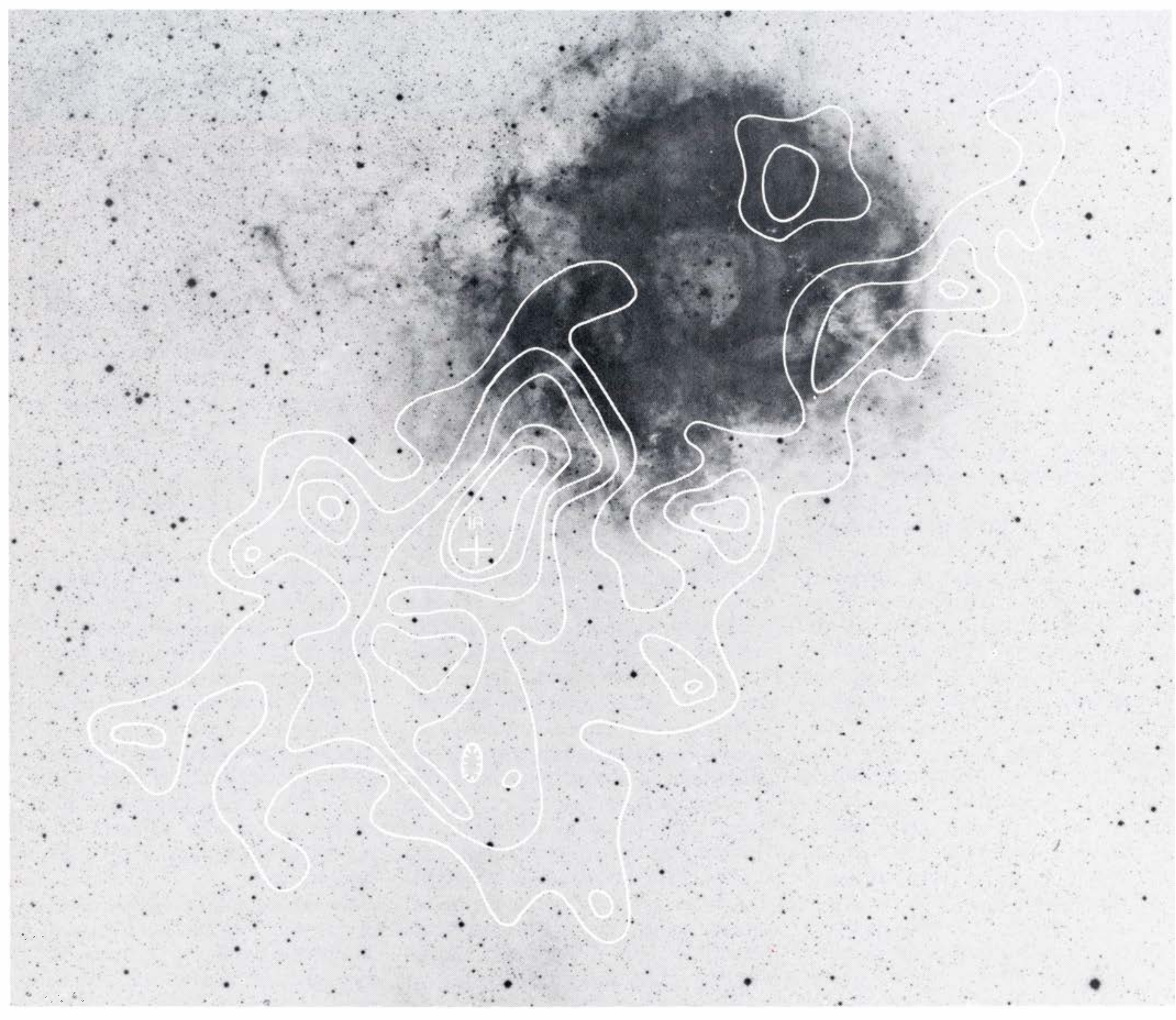
Que el medio interestelar del disco de la Vía Láctea incluía nubes de gas y polvo en las que se están condensando las estrellas nacientes era algo sabido ya hace 35 años. Después de todo, entre los objetos más sorprendentes de la galaxia se cuentan las nebulosas de emisión: brillantes nubes de átomos de hidrógeno ionizado que obtienen su energía de los grupos de estrellas gigantes dentro de ellas. Las estrellas emiten radiación a un ritmo cuya prodigiosa velocidad no permite una duración de las mismas más allá de escasas decenas de millones de años. Las estrellas son, pues, muy jóvenes. Pero con la mera observación óptica de las nubes resultaba difícil progresar en nuestro conocimiento sobre el proceso de formación de las estrellas.

El panorama ha cambiado. La radioastronomía ha revelado la presencia en las nubes de más de 50 clases de moléculas interestelares, desde moléculas de hidrógeno, las más ligeras y mil veces más corrientes, hasta cadenas de nueve carbonos, las más pesadas. Cada especie molecular emite radiación electromagnética a longitudes de onda características. Además, la astronomía de infrarrojo dispone ahora de medios para revelar las propias estrellas incipientes dentro de sus densas y oscuras nubes.

En casi todos los procesos en los que se cree que surge una nueva estrella, el paso inicial consiste en la formación de una concentración de materia –la llamaré un nódulo– dentro de una nube de átomos y moléculas interestelares (en su mayor parte, moléculas de hidrógeno) con alguna pequeña mezcla de polvo. Hay nubes que son enormes distribuciones de materia amontonada. Un complejo molecular gigante, por ejemplo, puede alcanzar una masa de varios cientos de miles de masas solares. Otras nubes son mucho más pequeñas; algunas de las llamadas glóbulos tienen masas de sólo 20 veces la solar.

Todos los nódulos comparten la tendencia a colapsar, principalmente bajo la influencia de su propia gravitación, con, quizá, presiones ocasionales desde fuera. Una presión exterior que puede ser muy importante es la ejercida sobre una nube y los nódulos incipientes dentro de ella, cuando una onda de densidad pasa a su través y comienza la formación de un brazo espiral. La onda puede muy bien acelerar cada uno de los procesos que describiré a continuación.

Las condiciones en que se hallan los aproximadamente 4000 complejos moleculares gigantes localizados dentro de los 13.000 parsecs del centro galáctico parecen hechas a medida para la formación de nuevas estrellas. Por un lado, cada complejo tiene abundancia de materia. La masa de un complejo típico es de varios cientos de miles de masas solares. Su diámetro mide unos 50 parsecs. Constituye el objeto más masivo de la galaxia (a menos que el centro galáctico posea un agujero negro supermasivo). Se cree que la masa de un complejo molecular gigante consta, casi exclusivamente, de moléculas de hidrógeno; a una temperatura de 20 grados K, la nube está demasiado fría para que las moléculas se disocien en átomos. La nube registra una temperatura también muy baja en su mayor parte, lo que impide que las moléculas de hidrógeno emitan cantidades detectables de radiación a sus longitudes de onda características. Debe deducirse, por tanto, la presencia de hidrógeno molecu-



UNA NEBULOSA BRILLANTE Y OTRA OSCURA que son vecinas en la constelación del Unicornio. Juntas integran un ejemplo del modo de formación de estrellas esquematizado en la ilustración precedente. La nebulosa brillante o de emisión, llamada la Roseta, es la nube oscura de la parte superior derecha de este negativo. Sus átomos de hidrógeno han sido ionizados por la radiación ultravioleta de un grupo de jóvenes estrellas supergigantes blancoazuladas. Se localizan varias estrellas de esas en lo que parece ser un agujero en el centro de la nebulosa. La radiación de la estrella en el mismo centro es muy intensa, y puede haber excavado el agujero. La nebulosa oscura es un complejo molecular gigante. Consta casi completamente de moléculas frías de hidrógeno. La totalidad de su extensión queda revelada mejor, sin embargo, por la radiación de las moléculas de monóxido de carbono a la longitud de onda de radio de 2,6 milímetros. Las líneas de la ilustración delinean la temperatura de antena, que indica la intensidad de la radiación del monóxido de carbono. La configuración de las líneas sugiere que la radiación ultravioleta de las estrellas supergigantes está erosionando el complejo. Al parecer, en el borde inferior izquierdo de la Roseta existe un frente de choque a lo largo del cual se está comprimiendo la materia del complejo. El frente es, presumiblemente, una zona donde se condensarán nuevas estrellas. Un máximo de la radiación del monóxido de carbono corresponde a una fuente de radiación infrarroja (*IR*), identificada ahora como una estrella recién nacida inmersa en gas y polvo. Tanto la nebulosa Roseta como el complejo molecular gigante se encuentran a 1600 parsecs del sistema solar. La ilustración representa los resultados del estudio de la interacción entre nebulosas realizado por Leo Blitz, de la Universidad de California en Berkeley, y Patrick Thaddeus, del Instituto Goddard de Estudios Espaciales.

lar. Como mejor se detecta la nube es a través de la emisión (a 2,6 milímetros) de la molécula siguiente en abundancia, el monóxido de carbono, y a través de la emisión de moléculas todavía menos corrientes, como el formaldehído.

En la formación de estrellas reviste interés otro componente de un complejo molecular gigante. Nos referimos al polvo, cuyas partículas son núcleos sobre los cuales se puede depositar el gas circundante. El polvo protege también a las estrellas nacientes de la radiación ultravioleta, que perturbaría la condensación. Se calcula un gramo de polvo por cada 100 o 200 gramos de hidrógeno molecular de un complejo molecular gigante.

Hay dos complejos vecinos de monóxido de carbono y polvo (y presumiblemente hidrógeno en abundancia) que han sido particularmente bien estudiados. Se trata del complejo de Orión (centrado en la brillante región de hidrógeno ionizado y estrellas gigantes recién nacidas llamada Gran Nebulosa de Orión) y el complejo de Ofiuco (que impide el paso de la luz procedente del centro de la galaxia). En el complejo de Orión, Becklin, Neugebauer y sus colaboradores, del Instituto de Tecnología de California, junto con Frank J. Low, de la Universidad de Arizona, han hallado, en longitudes de onda en el infrarrojo, indicaciones de la presencia de objetos concentrados que parecen ser intrínsecamente muy rojos. Cada uno de ellos puede significar una estrella recién nacida, todavía inmersa en una espesa envoltura de polvo que la radiación ultravioleta de la estrella no ha podido aún destruir por completo. La radiación calienta el polvo que, a su vez, radia entonces en el infrarrojo. La estrella puede haberse formado en primer lugar, cuando el enfriamiento de una pequeña parte del complejo redujo la presión resultante del calor de un gas. La región enfriada podría, así, haber empezado a colapsar bajo la influencia de su propia gravitación.

En el complejo de Ofiuco, Gary L. Grasdalen, de la Universidad de Wyoming, Stephen E. y Karen M. Strom, de la Universidad de Arizona, y Frederick J. Vrba, del Observatorio Naval de los Estados Unidos, han encontrado un grupo de 30 estrellas. No se habían detectado antes porque las ocultan 30 magnitudes, al menos, de absorción óptica: fue precisa una exploración mediante infrarrojo para hallarlas. Observaciones más recientes en una nube de polvo extremadamente densa, cerca de la estrella Rho de Ofiuco, llevadas a cabo por Charles J. Lada y Bruce A. Wilking, de la Universidad de Arizona, han revelado la presencia de 20 estrellas semejantes, cuya radiación óptica es atenuada por al menos 100 magnitudes. Estas estrellas se encuentran a sólo medio parsec de las anteriormente descubiertas.

Las indicaciones existentes hasta la fecha sugieren que los complejos moleculares gigantes originan espontáneamente estrellas, de masa no superior a unas cuantas veces la del Sol. En particular, parece que dan lugar a estrellas de las clases espectrales *B, A, F* y *G*. (El Sol es una estrella de clase *G*.) George H. Herbig y sus colaboradores, del Observatorio Lick, han encontrado, cerca de los bordes de complejos moleculares bien estudiados, algunos grupos de estrellas pequeñas, bastante tenues y nebulosas, que se suponen jóvenes. Casi todas ellas tienen un brillo que varía irregularmente. Se llaman estrellas T de Tauro. Se trata, quizá, de productos de la condensación interrumpida de nódulos. Tal vez se salieron del complejo. Se las observa frecuentemente en regiones donde la radiación ultravioleta de estrellas masivas recién nacidas ha barrido el oscuro gas y polvo de la nebulosa.

La formación de estrellas de otra clase parece ser típica de lugares donde una nebulosa de emisión, con estrellas supergigantes *O* y *B* inmersas en ella, se encuentra en la proximidad de un complejo molecular gigante. Bruce G. Elmegreen, de la Universidad de Columbia, y Lada han dado la más clara explicación de lo que sucede en tales casos. Según Elmegreen y Lada, las estrellas *O* y *B* emiten radiación ultravioleta cuya presión acumula gas frío y polvo en el borde exterior del complejo. El resultado es la condensación de protoestrellas allí. En ciertos lugares, en la nebulosa de Orión por ejemplo, el proceso parece estar avanzando de una manera secuencial; hay en ella un grupo de supergigantes *O* y *B* que se está debilitando tras una vida que dura algunas decenas de millones de años. La radiación de esas estrellas ha desencadenado la formación de una generación más joven de supergigantes *O* y *B;* las estrellas más jóvenes están ahora, a su vez, emitiendo radiación que horada lenta y persistentemente el complejo molecular gigante, mientras que un tercer conjunto de estrellas *O* y *B* supergigantes se halla, presumiblemente, a punto de formarse. Sigue sin saberse por qué el proceso da lugar a gigantes y supergigantes *O* y *B,* en vez de originar estrellas *B, A, F* y *G,* de menor tamaño, que se condensan espontáneamente en el complejo.

Volvamos a la clase de nubes oscuras

llamadas glóbulos. Se han identificado unos 200 de esos objetos dentro de 500 parsecs del Sol. Guardan una estrecha semejanza en sus propiedades. Son oscuras y bien definidas; aparecen casi circulares en la placa fotográfica. No cabe duda respecto de su esfericidad, casi completa. Su radio varía de 0,2 a 0,6 parsecs, su masa de 20 a 200 veces la del Sol y su temperatura interna de cinco a 15 grados Kelvin. Son impenetrables a la luz visible. Por otro lado, algunas imágenes registradas en longitudes de onda del infrarrojo próximo, bien fotográficamente o mediante técnicas electrónicas de reproducción de imágenes, muestran las estrellas detrás del glóbulo. La atenuación de la radiación infrarroja procedente de tales estrellas permite estimar el contenido de polvo del glóbulo.

Formación de estrellas en los glóbulos

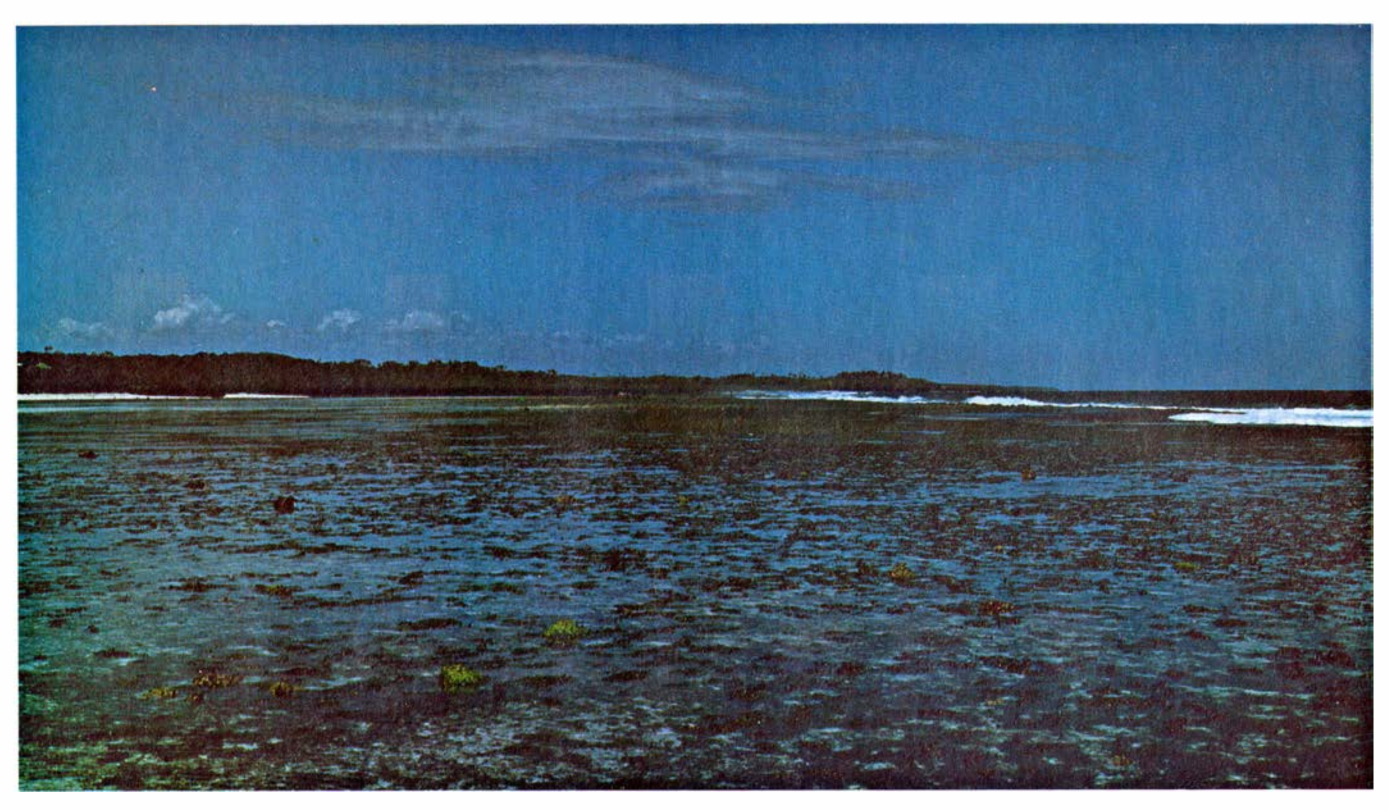
Radioobservaciones realizadas en 2,6 milímetros manifiestan que en los glóbulos abunda monóxido de carbono. Se han encontrado, igualmente, otras moléculas, en particular formaldehído y amoníaco. El glóbulo será, pues, en ese caso, una esférula pequeña y frecuentemente aislada, cuya obscuridad tiene mucho que ver con la de los complejos gigantes. Presumiblemente, el glóbulo consta, sobre todo, de hidrógeno molecular demasiado frío para emitir radiación detectable.

Los datos en la banda de radio muestran también que varios de los glóbulos se están colapsando bajo la influencia de su propia gravitación. El ritmo de colapso oscila en torno al medio kilómetro por segundo, lo que corresponde a medio parsec cada millón de años. Al medir medio parsec el radio de un glóbulo típico, el colapso tardará un millón de años aproximadamente en llevarse a término. El colapso vuelve a ser crucial en los modelos de formación de estrellas. En casi todos los modelos, el centro del glóbulo colapsa más deprisa que la periferia y se forma así un nódulo. El colapso convierte en energía cinética la energía potencial gravitatoria de la materia que cae hacia el interior. Con el tiempo, la energía del centro eleva la temperatura de la materia lo suficiente para que comience la fusión termonuclear. Esto señala el nacimiento de una estrella. Si es un ejemplar grande, emitirá suficiente radiación para barrer el gas y polvo envolvente.

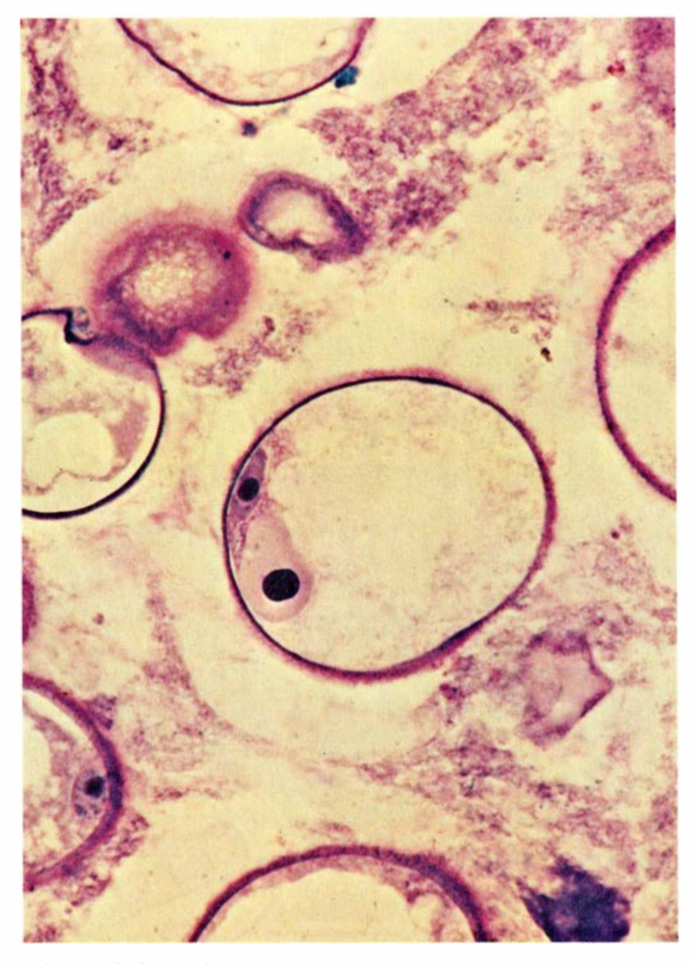
En resumen, un glóbulo debería invertir en torno al millón de años en dar nacimiento a una estrella única. Considerando la escala de tiempo de los procesos y estimando el número de glóbulos en la Vía Láctea, se llega a la conclusión de que los glóbulos podrían dar cuenta de la formación de 25.000 estrellas por cada millón de años, lo que viene a resultar una sexta parte del ritmo general al cual se forman las estrellas en la galaxia. (La formación de estrellas en complejos moleculares gigantes quizá sea un proceso más fecundo.) Pero en 1977, Richard Schwartz, observó lo que parecía ser una pareja de estrellas incipientes y nebulosas que estaban en fase de expulsión de un glóbulo muy oscuro, a unos 300 parsecs del sistema solar, en la parte del hemisferio sur celeste marcada por la constelación Vela. La creación de dos estrellas en lugar de una sorprende tanto como su expulsión del glóbulo. Uno se pregunta de dónde procede la energía para expulsarlas. Las dos estrellas están conectadas por una banda luminosa que se podría describir, en metáfora, como un cordón umbilical estelar.

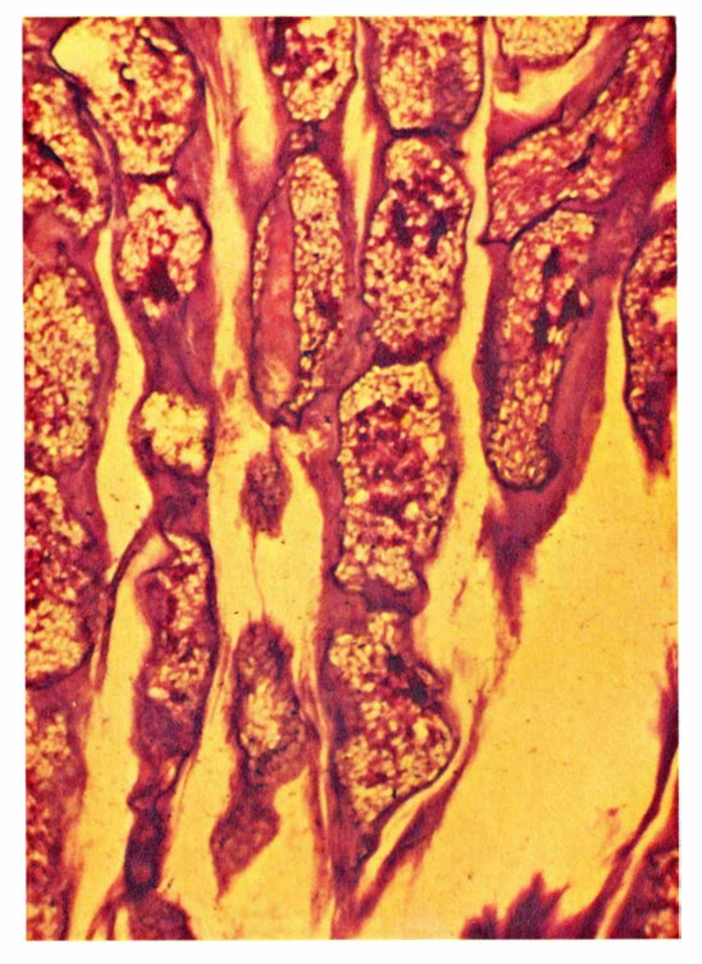
Querría dar un ejemplo final para ponderar cuánto queda por aprender. Los investigadores que intentan modelar los procesos a través de los cuales se forman las estrellas han supuesto, tácitamente, que se necesita la protección de una nube de polvo. Dentro de tal nube, la temperatura es baja y el proceso de deposición no queda perturbado. En particular, el polvo protege el interior de la nube de la alteración producida por la radiación ultravioleta procedente de otras fuentes externas. Considérense, empero, las Nubes de Magallanes, que están en la corona galáctica y desafían la generalización de que todo el contenido de la corona es antiguo.

En la Gran Nube de Magallanes hay una agrupación de unas 50 estrellas luminosas *O* y *B*, que reciben la denominación de Constelación I de Shapley. Es probable que no superen los 20 millones de años de edad. Sus velocidades son del orden de 10 kilómetros por segundo. De aquí que cada estrella se haya alejado no más de 200 parsecs del lugar donde nació. Dentro de un radio de 200 a 300 parsecs de donde están ahora las estrellas, las medidas a 21 centímetros han revelado cinco millones de masas solares de hidrógeno atómico neutro, y las medidas en longitudes de onda ópticas han puesto de manifiesto 60.000 masas solares de hidrógeno ionizado. Ahora bien, en esa región el cielo aparece transparente. Es probable que no aloje ningún complejo molecular gigante; tiene poco o ningún hidrógeno molecular, y carece de polvo cósmico. ¿Cómo se formaron allí las estrellas?



COLONIA DE HIERBA TORTUGA, una de las 50 especies de hierbas marinas polinizadas bajo el agua, sobre una plataforma erosionada por las olas en la costa de Kenia. Podemos apreciar la hierba tortuga en la fotografía merced a las condiciones de marea baja; con mareas altas quedaría sumergida.





DOS CLASES DE POLEN DE HIERBA MARINA aparecen en estas fotomicrografías. A la izquierda, granos inmaduros de la hierba tortuga, vistos en una sección de la antera, el órgano masculino de la planta. Como puede observarse en el grano central, se encuentran en la fase bicelular del desarrollo. Han comenzado a formarse tiras de limo (coloreadas en púrpura) que rodearán los granos de polen maduros. A la derecha aparecen las cintas lineales de *Halophila*. Cada cinta es un tubo mucilaginoso que contiene cuatro granos de polen. Los tubos aparecen en púrpura. Fotografías de Pettitt.

Polinización submarina

Las hierbas marinas, que florecen bajo la superficie de los océanos, deben liberar y capturar el polen en un medio subacuático. ¿Cuáles son las diferencias entre su sistema de reproducción y el de las plantas que habitan en tierra firme?

John Pettitt, Sophie Ducker y Bruce Knox

En el ciclo biológico de las plantas con flor, los acontecimientos de la polinización son de la mayor importancia. En el transcurso de su evolución, han desarrollado elaboradas estrategias para transferir el polen de una flor a otra; unas plantas dependen, para su polinización, del viento; otras, de abejas, murciélagos u otros animales. Sin embargo, estos mecanismos resultan inviables entre ciertos grupos de plantas con flores, por la sencilla razón de que florecen bajo el agua. Las hierbas marinas, que colonizan los hábitat costeros marinos de las regiones templadas y tropicales del mundo, constituyen uno de tales grupos.

¿Cómo se realiza la polinización bajo el agua? A primera vista, la respuesta parece sencilla: el polen se distribuye por los movimientos del agua, en particular por la acción de las olas. El método sería análogo al del polen transportado por el viento, pero un examen más detallado revela una fascinante complicación. Entre las plantas terrestres con flor, un elemento fundamental para que la reproducción tenga éxito es el "reconocimiento" químico entre ciertas moléculas de proteína o glicoproteína (proteínas con cadenas laterales de azúcar) del grano de polen, o tubo polínico, y determinadas moléculas de glicoproteína de la superficie receptora del estigma, u otras zonas del estilo, del pistilo, que es el órgano femenino de la flor. Se trata de moléculas hidrosolubles, una característica difícil de reconciliar con el reconocimiento del polen bajo el agua. ¿Poseen las hierbas marinas un sistema de reconocimiento diferente? Esta es la pregunta que motiva nuestra colaboración en el trabajo que aquí se resume.

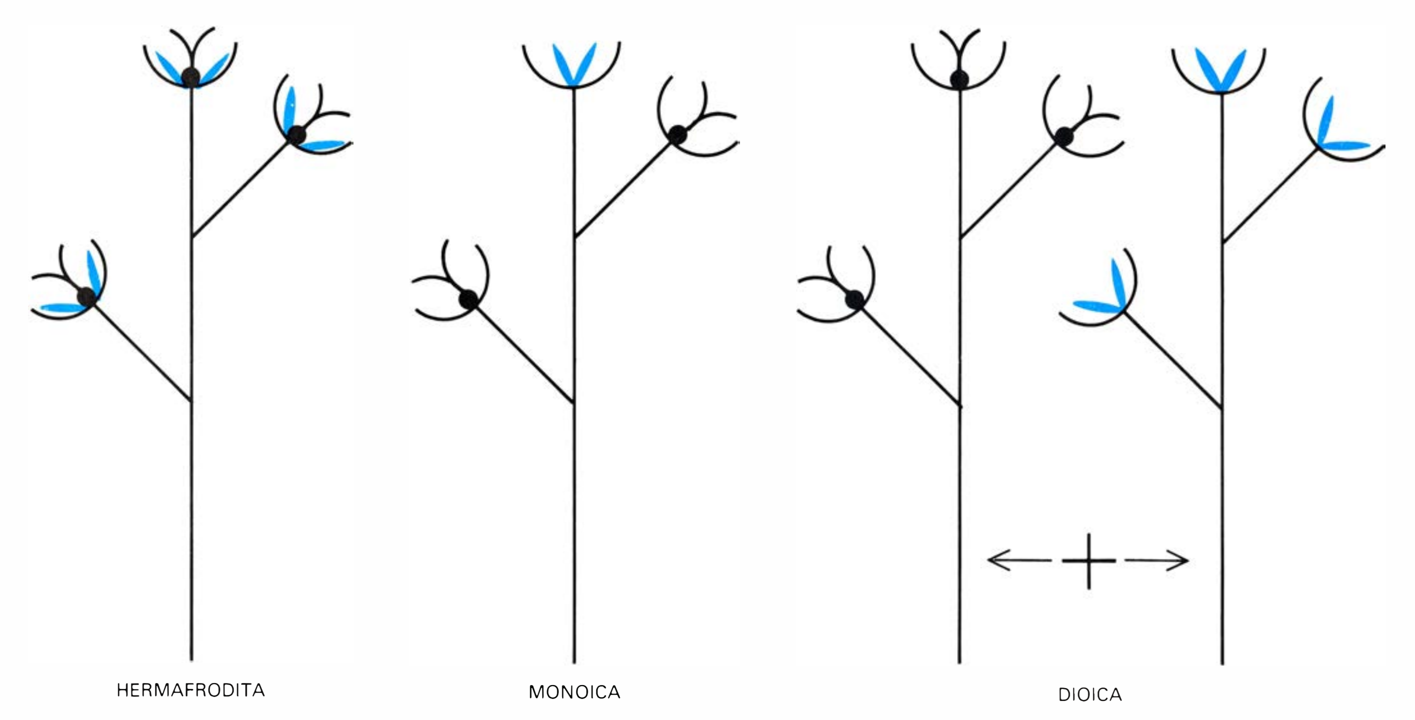
El nombre de hierba marina es puramente descriptivo, como lo es el de mala hierba marina que se da a las algas. Estrictamente hablando, las hierbas marinas no guardan relación con ninguna de las hierbas terrestres. Comprenden 12 géneros, con unas 50 especies. Son miembros del orden de plantas con flor que incluye la familia de las Potamogetonáceas y la de las Hydrocharitáceas. La zostera y la hierba de rompiente ("surfgrass") son ejemplos familiares en las regiones templadas. *Halophila* y talasia se encuentran ampliamente distribuidas en los trópicos.

Una amplia revisión realizada por Cornelis den Hartog, de la Universidad de Nijmegen, ha mostrado que la mayor concentración de hierbas marinas se registra en la región indomalaya. Allí pueden encontrarse ocho de los 12 géneros. Abundan también en la zona tropical del Nuevo Mundo, centrada en el Caribe. La distribución suscita muchas cuestiones en torno a los orígenes geográficos de las plantas, migraciones y especiación. Por ejemplo, la talasia, o hierba tortuga (del género *Thalassia*), consta de dos especies morfológicamente similares, que parecen estar íntimamente relacionadas y derivaron quizá de una población ancestral común. Una de las especies se encuentra en el Océano Indico y en el Pacífico occidental; la otra, en el Mar del Caribe y el Golfo de México. No se cruzan entre sí. ¿Se debe esta incapacidad para cruzarse a su separación geográfica o a un distanciamiento genético? Si se pusieran juntas, ¿podrían cruzarse de forma eficiente o exhibirían aislamiento reproductivo? La pregunta, sin respuesta todavía para la hierba tortuga, puede ya contestarse en el caso de otra hierba marina, la ninfa de mar (género *Amphibolis*). Este género también consta de dos especies, pero la distancia entre ellas es menor. Crecen juntas a lo largo de las costas del oeste y del sur de Australia. Al parecer, una barrera biológica impide el apareamiento interespecífico. Aunque los períodos de floración de ambas especies pueden coincidir, o al menos solaparse, no se observan descendientes híbridos.

Por lo que se refiere a la reproducción, las plantas con flor pueden agruparse en tres categorías. La inmensa mayoría son hermafroditas; sus flores contienen, pues, el órgano femenino, pistilo, y el masculino, anteras portadoras de polen. Algunas plantas son monoicas (del griego, "única casa"): la flor posee sólo pistilo o sólo anteras, aunque en una misma planta se encuentran ambos tipos de flores. Solamente el 3 o 4 por ciento de las plantas con flor son dioicas ("dos casas"): las flores femeninas y las masculinas se encuentran en plantas distintas. La incidencia de estos estados reproductores entre las hierbas marinas difiere considerablemente de la observada en plantas terrestres. Nada menos que nueve de los 12 géneros son dioicos.

Con independencia de su sexualidad, la secuencia de acontecimientos en la reproducción de las plantas con flores es básicamente la misma. Cuando la flor madura, la antera libera los granos de polen, que contienen gametos masculinos. Los granos deben alcanzar un pistilo maduro y adherirse a su estigma receptivo. El reconocimiento puede ocurrir en ese momento. Si el grano es compatible, germina y comienza a formar una estructura tubular. Este tubo polínico penetra la cutícula, crece a través del estilo del pistilo y acaba por llegar al ovario, donde se encuentra con el gameto femenino. Los gametos masculinos pasan a través del tubo polínico hacia el ovario, tiene lugar la fecundación y se inicia la formación de la semilla. Si el grano de polen no es compatible, la formación del tubo polínico se malogra, o, si se forma, no puede penetrar el estigma o no puede crecer a través del estilo. En cualquiera de estos tres casos, no se produce la unión entre el gameto masculino y el femenino.

Las plantas con flor dominan la vegetación terrestre. Las hierbas marinas muestran signos claros de descender de



LAS PLANTAS CON FLOR llevan generalmente sus órganos sexuales en una de estas tres disposiciones. La mayoría son hermafroditas, como en el dibujo de la izquierda; todas las flores poseen anteras (*color*) que producen polen y pistilos femeninos (*negro*) que capturan el polen liberado. Unas son monoicas: las flores masculinas (*color*) y femeninas (*negro*) crecen en la misma planta, tal y como se describe en el dibujo central. Otras, quizás el 4 por ciento del total de plantas con flores, son dioicas: cada planta lleva exclusivamente flores femeninas o masculinas. El 75 por ciento de las hierbas marinas son dioicas.

a *b* *c* *d* *e* *f* *g* *h*

PERIODO DE MICROSPORA

50 MICRAS

a *b* *c* *d* *e* *f* *g* *h*

PERIODO DE MICROSPORA

30 MICRAS

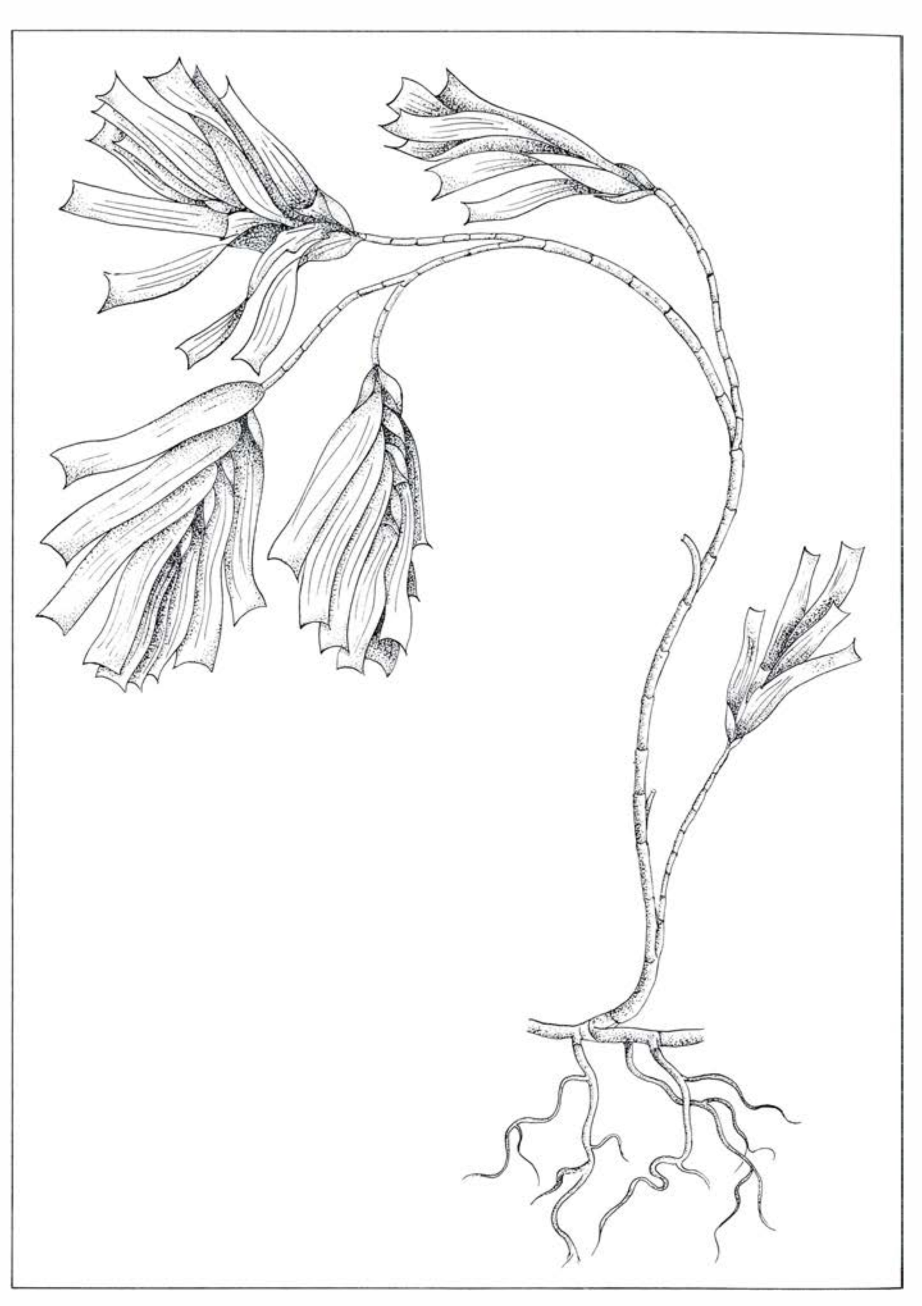
CINCO MILIMETROS

COMPARACION DEL DESARROLLO DEL POLEN en plantas terrestres con flor (*arriba*) y en plantas submarinas (*abajo*). La producción del grano de polen en el raigrás comienza con una tétrada de microsporas. Después de liberarse de la tétrada, cada microspora aumenta de tamaño; el período de microspora termina cinco estadios después; una división (*f*) forma una célula vegetativa (*negro*) y una célula progenitora espermática. Una segunda división (*h*) produce las dos células espermáticas del grano maduro; la maduración tarda dos semanas, a cuyo término el grano mide 50 micras de diámetro. La formación del grano de polen filamentoso de la ninfa de mar comienza también con una tétrada de microsporas. Tres estadios después, con el período de microspora ya terminado, la célula progenitora espermática (*color*) se ubica en un extremo del grano elongado (*d*). Posteriormente ocupa una posición central antes de dividirse en dos células espermáticas (*h*). la maduración dura 16 semanas; el grano mide entonces 30 micrometros de diámetro.

las plantas con flor terrestres. Parecen haber emigrado hacia el agua en épocas relativamente recientes. Se aprecian en ellas muchas diferencias morfológicas y estructurales con respecto a las plantas terrestres dotadas de flor, algunas de las cuales encierran sin duda un valor adaptativo en un hábitat submarino. No obstante, las hierbas marinas exhiben aún rasgos de su herencia terrestre, y la secuencia básica de los acontecimientos de su reproducción no se aparta sensiblemente de la de las plantas terrestres con flor.

Las adaptaciones florales de este tipo de plantas adquieren múltiples formas, aunque todas ellas deben estar encaminadas a una polinización eficaz, regulación de los niveles de variabilidad genética, nutrición del embrión y dispersión de las semillas en cantidad suficiente para el mantenimiento de la especie. La primera de estas medidas, la adecuada polinización, ha conducido al desarrollo de flores atrayentes en aquellas plantas cuya polinización depende de la atracción de animales. Las plantas polinizadas por el viento no necesitan de tales atractivos, por lo que sus estructuras florales pueden carecer, o carecen totalmente, de ellos.

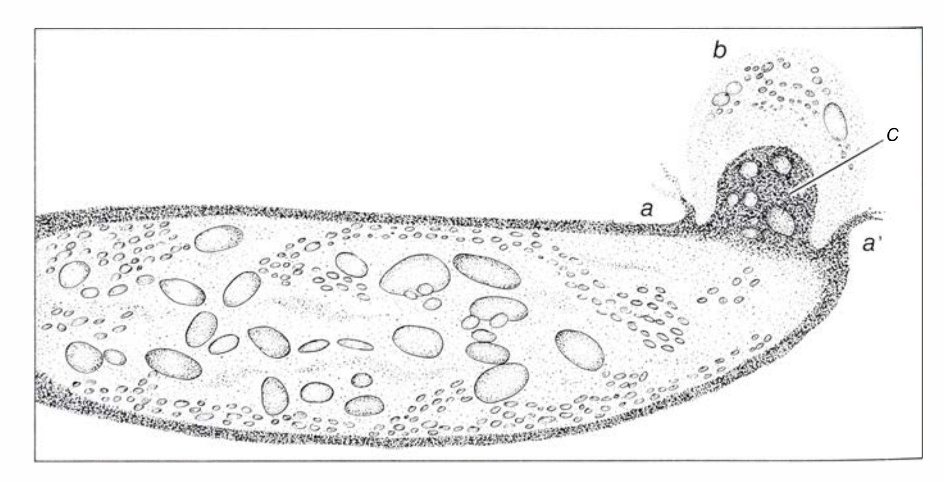
Lo mismo vale para las hierbas marinas. En estas plantas polinizadas por el agua, las estructuras florales que atraen a los insectos (pétalos coloreados de la corola o los sépalos del cáliz) son muy escasas, si es que llegan a estar presentes, y ofrecen mínima o nula ostentación. La talasia constituye un caso excepcional; sus flores, coloreadas, son unas estructuras petaliformes que presentan un color rojo púrpura brillante cuando brotan. Cualquiera que sea la razón para exhibir esas estructuras llamativas, ésta no es, desde luego, la de atraer a un potencial animal polinizador.



NINFA DE MAR, una hierba marina perteneciente al género *Amphibolis*. Utilizada por los autores en los experimentos de polinización en Australia, la planta alcanza entre 20 y 50 cm de altura con tallos erectos.

Con independencia de su carácter atractivo, la flor cumple dos funciones vitales: producción y recepción del polen. En las flores de las hierbas marinas, los órganos relacionados con estas funciones se hallan bien desarrollados. Los estambres maduros liberan granos de polen boyantes y los pistilos poseen estigmas receptivos para captar los granos flotantes.

En relación con la segunda condición para el éxito reproductivo, la regulación de la variabilidad genética, es significativo que la mayoría de los géneros de hierbas marinas sean dioicos. La segregación de las funciones sexuales entre individuos diferentes constituye una de las estrategias más eficaces para ase-



ABERTURA DEL GRANO DE POLEN EN LA NINFA DE MAR, que aparece tras la adhesión del grano de polen al estigma, flanqueada por fragmentos de la pared del polen (*a, a'*), disuelta ésta por acción enzimática. Emergiendo de la abertura se aprecia una burbuja mucilaginosa (*b*); debajo, una zona de citoplasma densamente coloreada (*c*) presagia el tubo polínico. Dibujo basado en una fotomicrografía.

gurar la variabilidad genética: refuerza el movimiento de polen entre plantas, favoreciendo la mayor frecuencia de variabilidad genética entre la descendencia. Por contra, la diecia no es un mecanismo de reproducción muy eficiente. La razón es que sólo la mitad de las plantas de una población, las femeninas, contribuyen a la producción de semillas, hecho que interesa a la última de las cuatro estipulaciones reproductivas. La elevada incidencia de plantas dioicas entre las hierbas marinas puede significar que la ganancia en supervivencia que resulta de la fecundación cruzada supera la desventaja que supone una menor producción de semillas.

La floración de las plantas terrestres responde a cambios en la longitud del día, temperatura y otros factores. ¿Cuáles son los factores que estimulan la iniciación y desarrollo de flores en las hierbas marinas? En los comienzos del siglo, el botánico sueco Nils Svedelius describió la floración de la hierba plana (género *Enhalus*), una hierba marina del Indo-Pacífico que había estudiado en Ceilán. Observó algo que ya sabían los pescadores locales desde hacía tiempo: su floración estaba sincronizada con el ciclo de las mareas. Si las mareas más altas del ciclo –las de primavera que siguen a la luna nueva y a la luna llena– ocurrían durante las horas de luz solar, la floración coincidía con ellas.

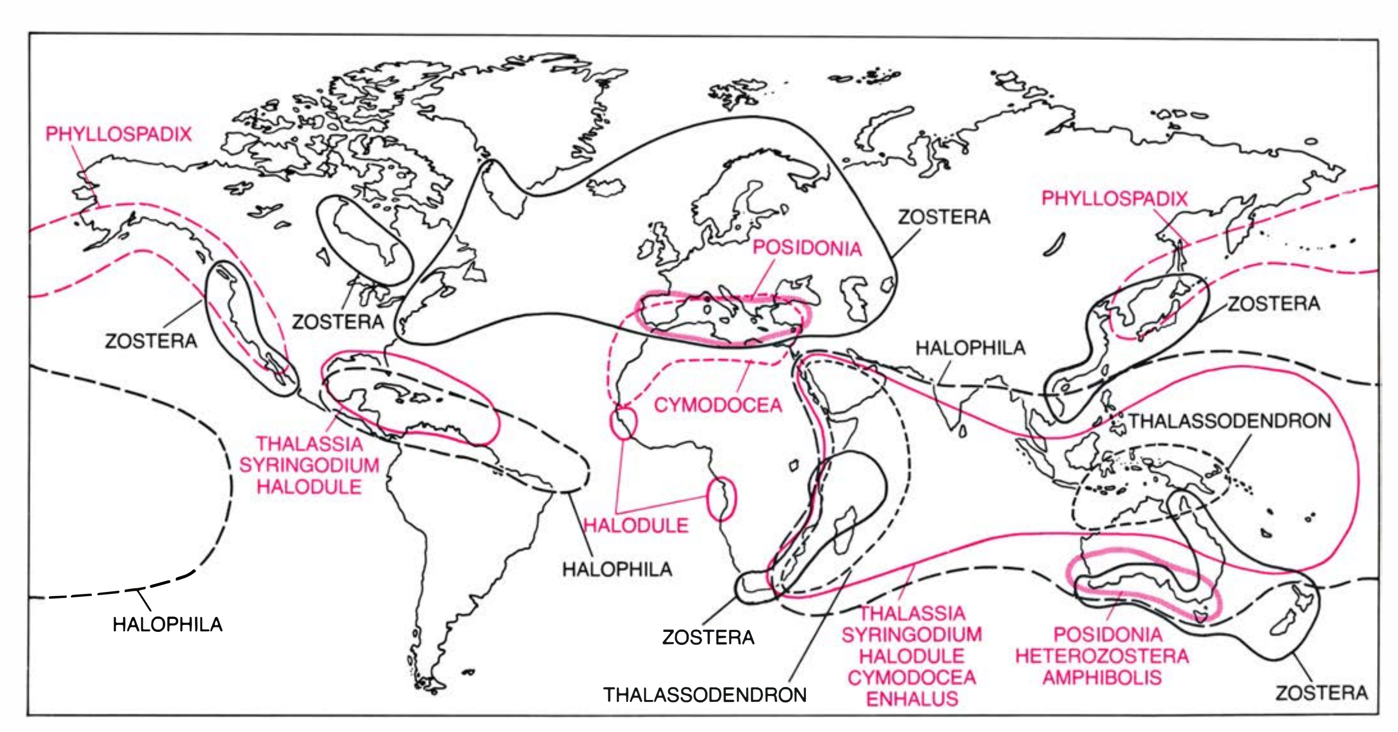
Hace cuatro años descubrimos una correlación similar en los hábitos reproductores de la hierba tortuga. Las colonias que estudiamos crecían en una plataforma superficial a lo largo de la costa de Kenia. Durante las mareas bajas, cuando la variación en el nivel del agua era mínima, las flores masculinas y femeninas comenzaban a emerger de las hojas que las recubrían. Al comienzo del período siguiente, las mareas de primavera, las plantas estaban en flor y se liberaba el polen. De esta forma, las plantas florecían cuando la circulación de agua a través de la plataforma alcanzaba su valor máximo.

Las flores se marchitaban horas después de la apertura. El período total de floración para todas las plantas de la población sólo duraba unos días. Las plantas que recogimos al final de las mareas de primavera estaban produciendo sus frutos o bien comenzando el desarrollo de flores que alcanzarían la madurez en las siguientes mareas de primavera. La liberación de semillas se mantenía sincronizada también con el ciclo de las mareas. Para una población de plantas que crecen en un hábitat costero, se trata de un mecanismo extremadamente eficaz para promover la distribución del polen y la dispersión de las semillas.

En otras hierbas marinas, la floración parece estar regulada más por las estaciones que por las mareas. Así, en las colonias de ninfa de mar que hay a lo largo de las costas de Victoria, Australia, las primeras estructuras florales se manifiestan a comienzos del invierno (mayo) y las flores maduran en el período que va desde principios a mediados del verano (desde mediados de noviembre hasta finales de diciembre). Las semillas de la ninfa de mar germinan mientras el fruto permanece aún unido a la planta madre. En las costas de Victoria, los frutos maduros, con sus plántulas, se esparcen en los meses de invierno, de mayo a agosto. El período de floración en las colonias de ninfa de mar de Australia occidental ocurre algo más tarde que en Victoria, lo que sugiere que la iniciación de las estructuras florales está regulada por la longitud del día.

El polen de las hierbas marinas ha atraído la atención de los naturalistas por lo menos desde finales del siglo XVIII, cuando Filippo Cavolini describió y dibujó en Nápoles el polen del género *Posidonia.* Cavolini observó que, con la apertura de las anteras, aparecía en el agua una gran acumulación de filamentos. Algunas décadas después, en 1826, el botánico francés Charles Gaudichaud-Beaupré observó, y también dibujó, el polen igualmente filiforme de la ninfa de mar en especímenes recolectados en la bahía de Shark, en Australia occidental. Algunas décadas más tarde, otro botánico francés, Edouard Bornet, describió el desarrollo del polen en colonias de la hierba marina *Cymodocea* en aguas mediterráneas.

A comienzos del siglo XX, el botánico sueco Otto Rosenberg fijó, seccionó y tiñó una serie de anteras de la hierba *Zostera*. También describió el proceso de diferenciación del polen desde su

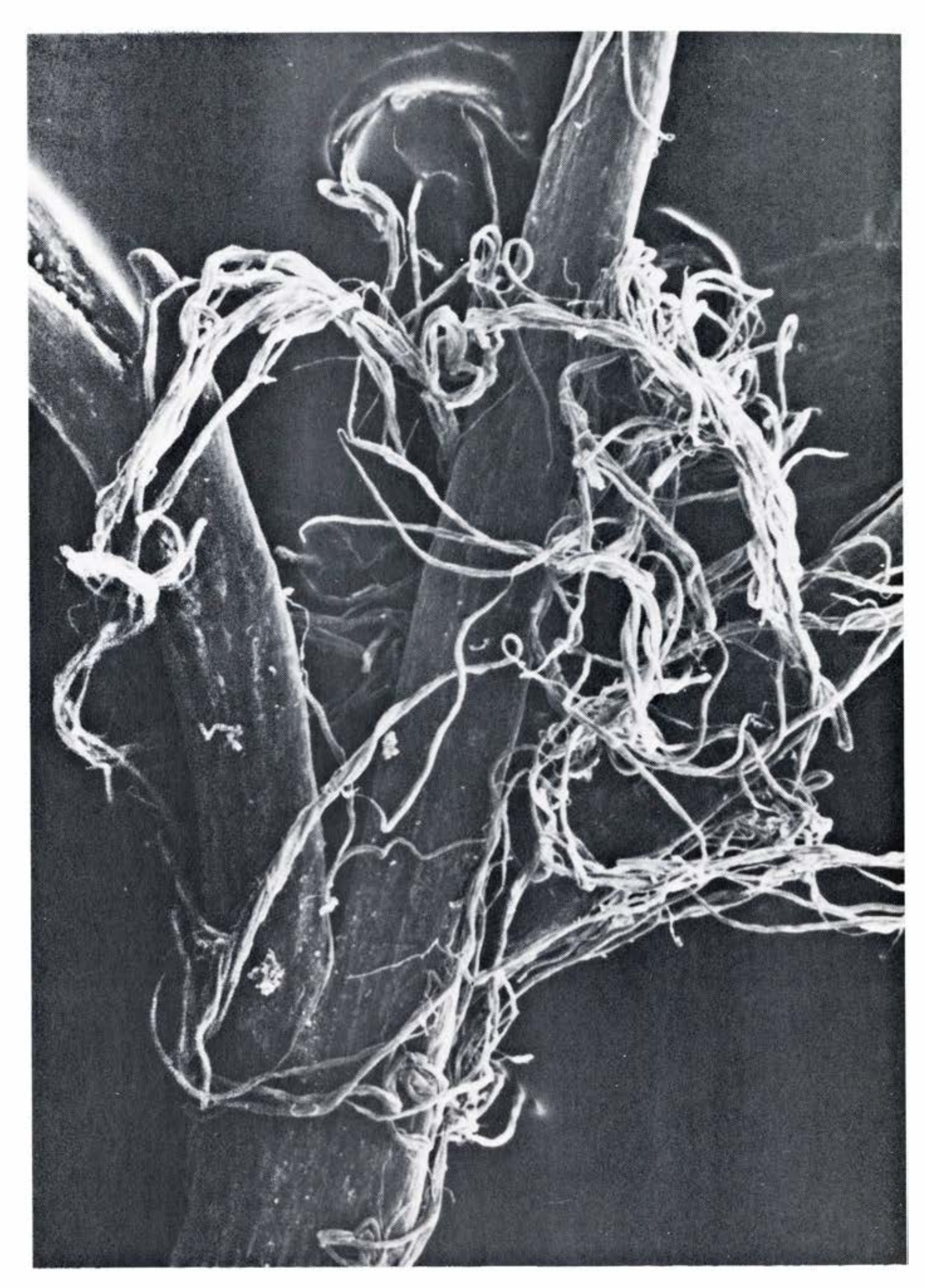


LAS PRINCIPALES CONCENTRACIONES de géneros de hierbas marinas se localizan en los trópicos de la región indomalaya con 12 géneros y en el Caribe, con cuatro. La cosmopolita *Zostera* ha llegado en su colonización hasta el mar Aral. *Posidonia* y *Cymodocea* pueden apreciarse en el Mediterráneo.

origen, en el momento de la meiosis (división celular que reduce a la mitad el número de cromosomas), hasta su madurez. En este género el polen es también filiforme. El proceso de diferenciación dura algunas semanas y los filamentos de polen maduro alcanzan una longitud de aproximadamente tres milímetros. Las células espermáticas, los gametos masculinos portados por el polen, no miden más de cinco o diez micrometros (milésimas de milímetro) de diámetro.

En los 80 años transcurridos desde los trabajos de Rosenberg con el polen de zostera, otros estudios similares han revelado las características generales del desarrollo del polen en distintas hierbas marinas. Mediante procedimientos adecuados para la detección de componentes celulares específicos, no sólo ha sido posible poner de manifiesto con considerable detalle la estructura y composición de los granos de polen, sino también demostrar el papel de estos componentes en la reproducción.

Los granos de polen de las hierbas marinas presentan ciertas características especiales que no parecen ser otra cosa que derivaciones adaptativas a partir del polen de sus antepasados terrestres. Para apreciar la importancia de estos cambios adaptativos es útil considerar primero la polinización terrestre. El grano de polen de una planta con flores terrestre suele ser esférico, de 30 a 60 micrometros de diámetro en el momento de su liberación de la antera. Según la especie vegetal, el grano maduro constará de dos o tres células. Si hay sólo dos células, una es vegetativa y la otra es el progenitor del gameto, que se divide y forma dos células espermáticas tras la germinación del polen. En un grano de polen con tres células, una es la vegetativa y las otras dos son las células espermáticas.

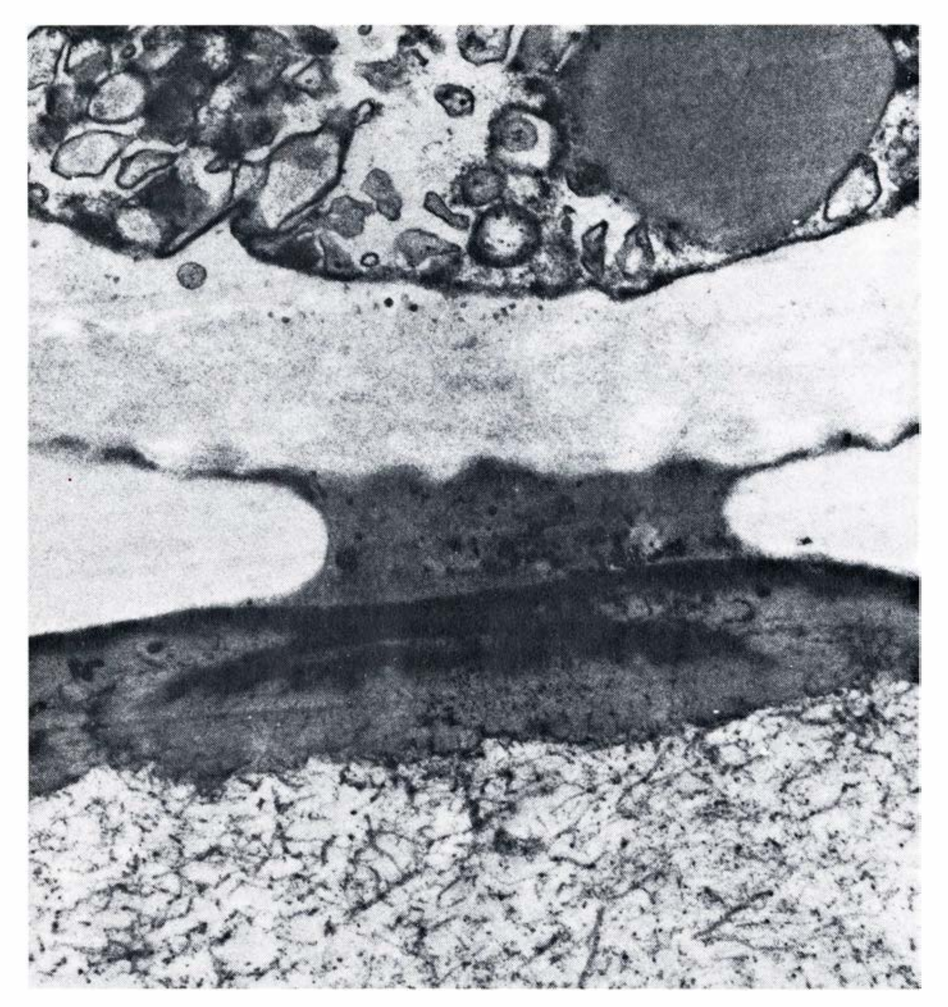


ESTIGMAS DE LA NINFA DE MAR, cuyas estructuras ramificadas se observan en esta microfotografía de barrido. Los estigmas han capturado granos de polen filamentosos, liberados por la planta masculina y dirigidos hacia las flores femeninas por los experimentadores. El aumento es de unos 20 diámetros.

Las paredes del grano de polen que rodean las células espermáticas constan de dos capas, que difieren en su origen, estructura y composición química. La capa externa está compuesta por esporopolenina, que resiste la degradación y presenta a menudo un elaborado esculpido. La capa interna no muestra ningún tipo de esculpido y está formada por finas fibrillas de celulosa incluidas en una matriz de polisacáridos. Los materiales para la construcción de la capa externa se originan en el tapete, el tejido nutritivo de la antera; la capa interna es un producto del protoplasma del grano de polen.

En la pared de dos capas del grano de polen se encuentran frecuentemente aberturas, o poros, que varían en forma, número y posición de acuerdo con las especies. Cuando germina el polen, el tubo polínico emerge a través de una de las aberturas. En la capa interna del grano se encuentran ciertos enzimas, como la fosfatasa ácida, cuya concentración es particularmente elevada cerca de las aberturas. En granos de polen que no tienen aberturas, el tubo polínico puede emerger a través de la pared del grano por cualquier parte; en este caso, los enzimas se hallan uniformemente repartidos por la capa interna del grano de polen. En cualquier caso, el resultado es que la punta del tubo polínico arrastra en su emergencia a los enzimas. Estos pueden también ser transferios desde las células del tapete de la antera hasta la capa externa de la pared del grano en una fase posterior del desarrollo del polen; tal transferencia fue demostrada a principios de la década de los setenta por John Heslop-Harrison, que estaba entonces en los Reales Jardines Botánicos de Kew, y uno de nosotros (Knox). Entre los enzimas transferidos se observó una esterasa, que ahora se la considera proteína característica de la capa externa.

Los granos de polen de las hierbas marinas son de tres tipos básicos: filiformes, esféricos y elipsoidales. Los granos filiformes son característicos de la zostera, la ninfa de mar, el género *Heterozostera* y el *Thalassodendron.*



ESTIGMA Y GRANO DE POLEN de la ninfa de mar, aumentados 15.000 diámetros en esta micrografía obtenida con el microscopio electrónico de transmisión. El eslabón oscuro entre la película superficial del estigma, en la parte inferior, y la del grano de polen, en la superior, establece el contacto entre ambas.

Estos cuerpos filamentosos, que miden hasta cinco milímetros de longitud y sólo 10 o 30 micrometros de diámetro, se distinguen, en su morfología, del polen de las plantas terrestres. En el grano de polen falta la capa externa de esporopolenina; la pared del polen está formada exclusivamente por microfibrillas de celulosa incrustadas en una matriz de polisacáridos. Cuando están frescos, los granos de polen filiformes dan reacción positiva para la fosfatasa ácida (el enzima que se encuentra en la capa interna de las paredes del grano de polen de las plantas terrestres con flor).

La superficie de los granos de polen filiformes está cubierta por una fina película de un material formado por carbohidratos y proteína. Incluida en esta fina película hallamos el enzima esterasa (el producto de las células del tapete que se localiza en la capa externa del polen de las plantas terrestres); este enzima se deposita en la fase final de la maduración del polen. De esta forma, aun cuando existen diferencias morfológicas entre el polen filiforme de las hierbas marinas y el polen de las plantas terrestres con flor, existen también semejanzas estructurales y bioquímicas.

Dos de las hierbas marinas tropicales y subtropicales, hierba tortuga y hierba plana (esta última del género *Enhalus*) producen granos de polen esféricos. Los de la hierba tortuga miden 100 micrometros de diámetro y, los de la hierba plana, 150 micrometros; ambos, sin aberturas. Su polen se parece al de las plantas terrestres en que posee, diferenciadas, capa interna y capa externa. En la capa interna descubrimos fosfatasa ácida; en la delgada película que cubre la capa externa, esterasa.

En el curso de nuestros trabajos en el Océano Indico observamos que los granos de polen de la hierba tortuga, al ser liberados al mar, quedaban embebidos en una especie de limo viscoso e insoluble; la masa del polen se asemejaba a la freza de los anfibios, aunque el polen es mucho más pequeño. Carl Julius Fritsche, químico alemán del siglo XIX, dedicó su atención a la presencia de este limo en las plantas acuáticas. Nos intrigaba la posibilidad de que esta sustancia insoluble interviniera en la polinización de la hierba tortuga, quizá pegando los granos de polen a los estigmas receptivos. Siguiendo el desarrollo del polen en la hierba tortuga, llegamos hasta el origen del limo en cuestión. El protoplasma de la célula tapetal invadía la cavidad de la antera y se convertía en limo en la fase final del desarrollo del polen.

Para determinar la composición del limo, lo separamos del polen por centrifugación. Los análisis revelaron un contenido de un 5 por ciento de carbohidratos. Los azúcares componentes del carbohidrato se identificaron por cromatografía. Más de la mitad del azúcar era manosa; el resto, xilosa, galactosa, glucosa, arabinosa y ramnosa. Mediante electroforesis se determinaron, en razón de su tamaño molecular, las proteínas y glicoproteínas del limo. Se encontró un único componente glicoproteico.

El examen microscópico del polen centrifugado revelaba que, después de la centrifugación, una delgada película de limo seguía recubriendo cada grano. Este descubrimiento sugería que aún después de haberse roto la masa del polen por acción de las olas, algunos restos de limo permanecerían adheridos a los distintos granos de polen y servirían de adhesivo para pegarlos a la superficie del estigma.

El tercer tipo de polen de las hierbas marinas es el producido por el género *Halophila.* Estos granos elipsoidales miden unos 40 por 80 micrometros y carecen de aberturas. El polen se libera de las anteras en forma de cintas o cordeles; cada cinta es un tubo mucilaginoso que contiene cuatro granos de polen. Los cuatro granos son células hermanas que se originan a partir de la misma célula madre en la meiosis. La pared del grano no tiene capa externa y la interna ostenta la constitución típica: microfibrillas en una matriz polisacarídica. Posee fosfatasa ácida. El tubo mucilaginoso, de construcción similar, está protegido por una delgada película superficial de proteínas, esterasa incluida, derivadas del tapete. La presencia de esta película superficial sugiere cierto protagonismo en el reconocimiento y unión del polen.

La captura y retención de los granos de polen depende también de la superficie receptiva del estigma de la hierba marina. En algunos géneros, la superficie es lisa o rugosa; en otros, presenta unos pequeños salientes. Ambas clases de superficies están cubiertas por una cutícula delgada que, a su vez, lleva una secreción superficial.

Cabe comparar las secreciones del estigma con las de ciertas plantas te-

rrestres. Uno de nosotros (Knox) trabajando con John y Yolande Heslop-Harrison y Ole Mattsson, en los Reales Jardines Botánicos, determinó varias características químicas de la capa superficial del estigma en plantas terrestres con flores; así, la presencia de esterasa y afinidad para la concanavalina *A*, una proteína encontrada en la judía de caballo (*Canavalia*), que se encuadra en un tipo de sustancias denominadas lectinas. Se sabe que la concanavalina *A* se combina con varios de los azúcares encontrados en polisacáridos y glicoproteínas, manosa y glucosa incluidas. Recientemente, Adrienne Clarke y Paul Gleeson, de la Universidad de Melbourne, han encontrado que la lectina animal tridacnina, producida por el eulamelibranquio *Tridacna maxima*, también se une con la superficie del estigma de las plantas terrestres con flores. Se sabe que la tridacnina tiene afinidad por ciertos azúcares. La secreción de la superficie del estigma de las plantas terrestres con flores incluye un componente azucarado y está relacionada químicamente con las gomas vegetales. Su función puede ser la de capturar y retener granos de polen, iniciando de esta manera el proceso de germinación.

Las secreciones superficiales de los estigmas de todas las hierbas marinas que hemos examinado se parecen a las de las plantas terrestres en la presencia de esterasa. Es más, la secreción de los estigmas de la hierba tortuga presenta afinidad por la concanavalina *A*. Existe, sin embargo, una diferencia significativa. Las secreciones de los estigmas de las plantas terrestres son hidrosolubles, en tanto que las de las hierbas marinas, una vez formadas, no se dispersan en el agua. La secreción retiene su integridad estructural y, como los experimentos de polinización han mostrado, cuando el apareamiento es compatible la secreción es sensible al contacto del polen.

El examen de los estigmas de las hierbas marinas revela una correlación morfológica entre las estructuras del polen y del estigma, que podría muy bien estar relacionada con la captura del polen, un proceso que depende de la probabilidad de colisión en un ambiente marino. Los estigmas cubiertos con pequeñas proyecciones pueden capturar polen esférico (polen cubierto con limo, como en la hierba tortuga y en la hierba plana) o pueden capturar cintas de polen con tubos mucilaginosos, como en *Halophila*. Los estigmas que carecen de tales proyecciones atrapan, invariablemente, granos filiformes, característicos de la zostera y de la ninfa de mar.

Entre las plantas terrestres, las modificaciones de la superficie del estigma afectan drásticamente a la polinización. Por ejemplo, uno de nosotros (Knox), trabajando con Clarke en la Universidad de Melbourne, encontró que, cuando la concanavalina *A* se aplicaba al estigma de *Gladiolus*, se desorganizaba la polinización compatible. El estigma perdía su capacidad para retener los granos de polen de *Gladiolus* y no se producía la penetración del tubo polínico en el estigma.

El experimento sugiere, primero, que bajo circunstancias normales la secreción superficial del estigma de las plantas terrestres constituye un adherente muy efectivo para el polen. También sugiere que los lugares de la superficie del estigma que se unen a la lectina desempeñan un papel crucial en el reconocimiento del grano de polen. Se han obtenido pruebas adicionales en los experimentos realizados por Heslop-Harrison y Hugh Dickinson con *Agrostemma* y coles (*Brassica*) en el laboratorio de Dickinson de la Universidad de Reading. Mediante eliminación o modificación de la secreción de los estigmas, con enzimas, lograron frustrar el proceso normal de la polinización.

Llegamos así a la cuestión siguiente: ¿Es diferente la captura, retención y germinación del polen en el dominio marino? En nuestro laboratorio de la Universidad de Melbourne preparamos cultivos de ninfa de mar al objeto de estudiar el proceso de polinización, recolectando plantas masculinas y femeninas procedentes de colonias de Bass Strait y transfiriéndolas a acuarios con agua de mar. Recogimos el polen liberado por las plantas masculinas y lo dirigimos hacia los estigmas sumergidos de las plantas femeninas. La mayoría del polen no dio con ningún estigma, pero todos los granos que establecieron contacto quedaron adheridos y retenidos inmediatamente.

A ciertos intervalos, se retiraban las flores polinizadas de las plantas femeninas y se preparaban los estigmas para el examen microscópico. Se cortaron secciones delgadas de la interfase polen-estigma y se examinaron mediante microscopía electrónica. El análisis demostró que el grano estaba unido con firmeza al estigma por un menisco de adhesión. La unión, evidentemente, se formaba en el momento del contacto por coalescencia de materiales de las dos superficies. Tras el contacto con la secreción del estigma, la película de carbohidrato y proteína de la superficie del grano de polen se combinaba con ella, formando una especie de ligamento. La acción de ligar recordaba a la producida por ciertos adhesivos industriales que constan de dos componentes y que se fijan después de mezclarse.

Tal adhesivo de dos componentes le proporciona a la ninfa de mar el medio de discriminar entre el polen de su misma especie y el polen extraño. Si la adhesión depende de la interacción de macromoléculas específicas en las superficies de contacto, el polen extraño que disponga de una película superficial con características moleculares no adecuadas no se unirá al estigma.

Un experimento anterior nos había demostrado que el polen viable de la ninfa de mar era incapaz de germinar cuando se incubaba en recipientes llenos de agua de mar. El descubrimiento tiene una especial relevancia a la hora de interpretar los experimentos de polinización, ya que puede concluirse que es necesario el contacto con el estigma para estimular la germinación del polen. ¿Cuál es la forma que toma este estímulo? En las plantas terrestres con flores el estímulo es la hidratación: cuando los granos de polen deshidratados se posan, toman agua del estigma. El grano humedecido pasa de un estado de latencia a otro activo, durante el cual emerge el tubo polínico.

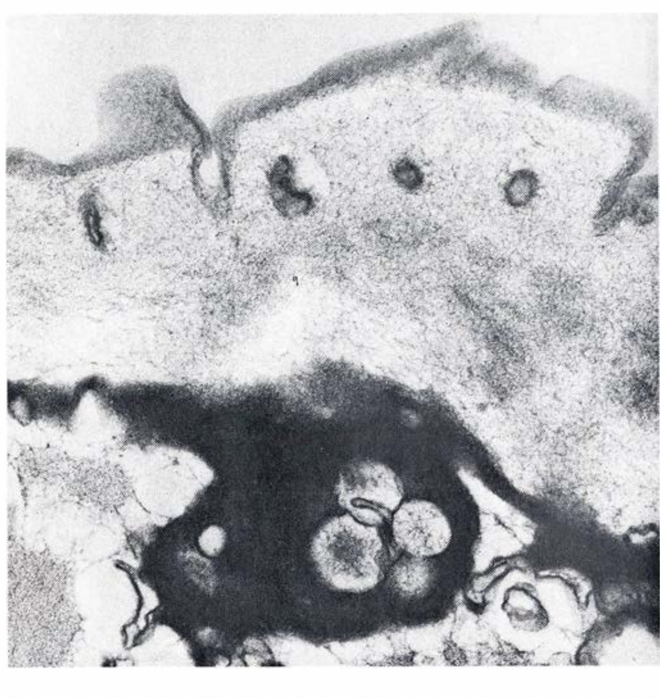
La hipótesis de que la hidratación desencadena de igual forma la germinación en medios submarinos presenta no pocas dificultades. Resulta dudoso, por ejemplo, que el grado de deshidratación del grano de polen sumergido pueda aproximarse al que se da en el polen terrestre. De todas formas, el contenido hídrico del grano de polen de una hierba marina puede ser inferior al del estigma correspondiente.

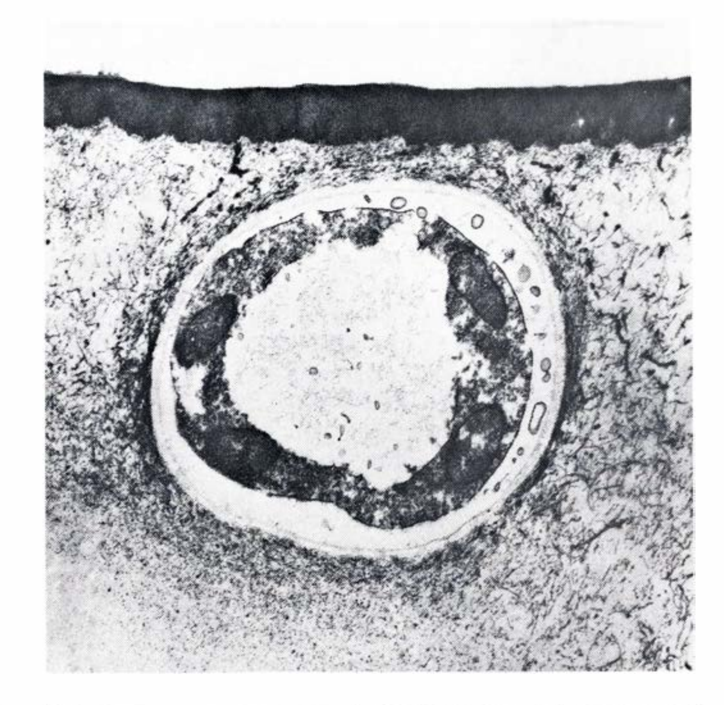
Una posible solución a estas dificultades se desprende del trabajo de A. W. A. M. de Cock, de la Universidad de Nijmegen. Investigando con polen de zostera, de Cock observó que el desarrollo del tubo polínico podía mejorarse si se añadía un extracto de pistilos triturados de la misma planta, junto con azúcares, al agua de mar que se utilizaba como medio. Sugirió que una molécula orgánica del extracto de pistilo (molécula que en circunstancias normales podría difundir del estigma) activaría el desarrollo del tubo polínico. Tal conjetura presenta implicaciones muy atractivas. Si tal molécula participa en la promoción del crecimiento del polen de las mismas especies, también podría intervenir en el reconocimiento de especies en el momento de aparearse.

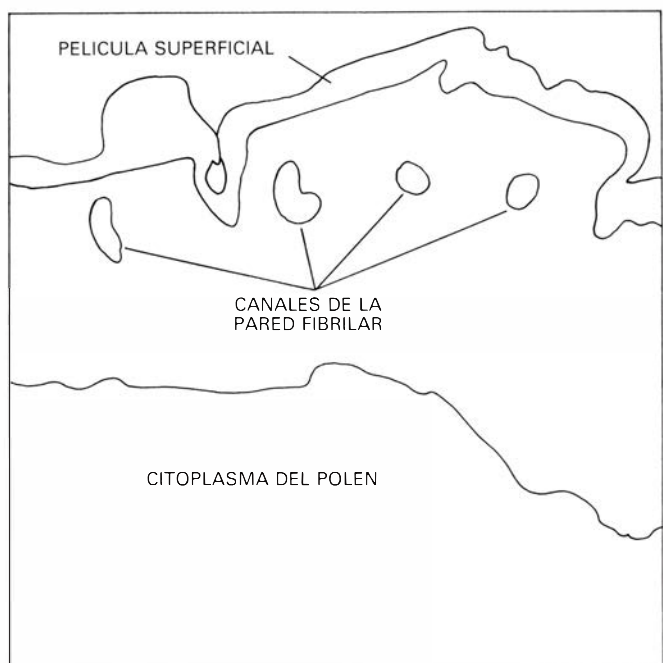
Puesto que algunas moléculas procedentes del estigma pasan al grano de polen para desencadenar su germinación, deberán atravesar la pared celular del grano. La realización del proceso dependerá del tamaño de las moléculas. Las sustancias de bajo peso molecular podrían atravesar la pared por difusión, mientras que conductos especiales podrían facilitar el paso de moléculas mayores. Hemos descubierto recientemente tales conductos. Atraviesan la pared celular de los granos de polen del género *Heterozostera* y miden unos 100 nanómetros (milésimas de micrómetro) de diámetro.

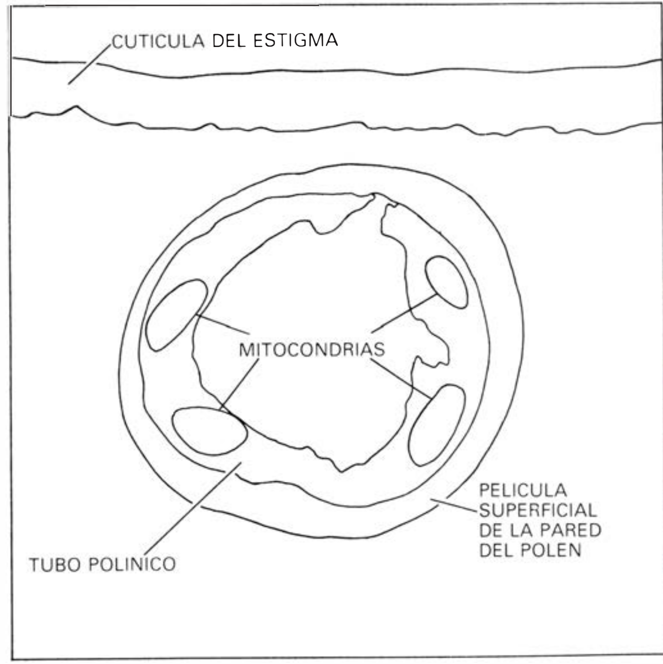
Cuando observamos la emergencia del tubo polínico en la ninfa de mar nos encontramos con algo inesperado. En el polen de las plantas terrestres con flores, la capa interna del grano se extiende para formar la primera parte del tubo polínico. Sin embargo, en el grano de polen de la ninfa de mar, se disuelve una región de la pared por acción enzimática, según parece. El microscopio permite detectar la incipiente abertura en un área pequeña de la pared del grano que pierde su afinidad por ciertos colorantes. Esta zona queda muy cerca del punto de unión entre el grano de polen y el estigma, pero nunca en el punto mismo. Ello indica que los enzimas responsables de la disolución de la pared se originan en el grano de polen, no en el estigma.

En el polen de la ninfa de mar, la emergencia de una burbuja mucilaginosa a través de la abertura de la pared









LAS MICROESTRUCTURAS DE GERMINACION, que aparecen en esta micrografía al microscopio electrónico de transmisión, están aumentadas 50.000 y 18.000 diámetros, respectivamente. La microfotografía de la izquierda muestra los pequeños canales de la pared fibrilar del polen de *Heterozostera*. Pueden servir de conductos para el paso, del estigma al citoplasma del polen, de sustancias que podrían estimular la germinación. También se distingue la delgada película superficial de la pared del grano de polen. La micrografía de la derecha muestra una sección transversal de un tubo polínico creciendo hacia el tejido del estilo de un pistilo después de la penetración de la pared de un estigma receptivo. Sobre el tubo polínico se observa la densa cutícula del estigma que el tubo ha penetrado en algún punto cercano. Las estructuras principales aparecen identificadas en los dibujos inferiores.

del grano de polen nos indica que ha comenzado la germinación. Bajo la burbuja hay una zona del citoplasma del polen densamente coloreado y delimitada por la membrana externa del citoplasma. El citoplasma pasa a través de la abertura para establecer un tubo polínico adecuado. Aunque se desconoce cómo se forman los materiales del tubo polínico, la síntesis de los componentes polisacarídicos de la pared parece darse en la parte frontal del citoplasma que emerge.

El desarrollo del tubo polínico en la ninfa de mar difiere del de las plantas terrestres con flores. En todos los casos conocidos de tierra firme, el desarrollo del tubo mantiene una continuidad estructural con la capa interna microfibrilar de la pared del grano de polen. Se ha descubierto recientemente una excepción a esa regla en los estados iniciales de la germinación del polen en el aguacate (género *Persea*). Margaret Sedgley, de la Australian Commonwealth Scientific and Industrial Research Organization de Adelaida, ha observado que los tubos polínicos de esta planta constan sólo de membrana citoplasmática externa; la síntesis de materiales de las paredes del tubo se retrasa hasta momentos después de que éste y el estigma entren en contacto. Esta aparente similitud de desarrollo entre una hierba marina y el aguacate podría ser fortuita; sin embargo, la existencia de la misma clase de fenómenos en especies tan diferentes incita a buscar una explicación funcional.

Cuando comenzaron nuestras investigaciones conjuntas, no se había observado la interacción del grano de polen con el estigma en ninguna de las hierbas marinas. Era obvio que nuestros estudios deberían revelar algunas divergencias estructurales y fisiológicas con las plantas terrestres, pero tampoco era disparatado esperar que los principios generales de la interacción polen-estigma fueran similares en el mar y en tierra firme. Como se ha demostrado en este trabajo, tal es el caso.

¿Cuáles son entonces los requisitos fisiológicos específicos para la polinización submarina? Para responder a esta pregunta habrá de conocerse la función precisa del mucílago del polen. ¿Sirve simplemente de adhesivo? ¿Regula el equilibrio hídrico del grano de polen? ¿O bien cumple ambos propósitos y evita, como la botánica inglesa Agnes Arber sugirió hace más de medio siglo, que las sustancias solubles (que podrían incluir las moléculas de reconocimiento) se difundan en el agua?

Percepción anortoscópica

Si a través de una rendija se van mostrando sucesivamente las secciones de una figura, ésta puede percibirse en su integridad. ¿Se basa tal efecto en la impresión de la imagen en la retina, o acaso en procesos psicológicos más profundos?

Irvin Rock

La cuestión del cómo percibimos visualmente la forma de las cosas, aunque es un tema ya viejo en psicología experimental, sigue sin resolverse. En algunos aspectos, el ojo es como una máquina fotográfica, puesto que todo objeto de determinada forma da origen a una imagen de la misma forma en la superficie de la retina. La simple analogía se viene, empero, abajo en situaciones en las que la presencia de esa imagen no es ni suficiente ni necesaria para explicar la forma percibida. Los investigadores han ideado pacientemente diversas técnicas de laboratorio para adquirir más conocimientos sobre los procesos perceptuales. Mis colegas y yo, en el Instituto de Estudios Cognitivos de la Universidad de Rutgers, hemos practicado con preferencia un método que requiere mirar una tras otra las secciones de una figura a través de una estrecha rendija. En reconocimiento de lo insólito de este dispositivo experimental, hemos adoptado el decimonónico término de 'anortoscópica' ("anormalmente mirada") para caracterizar la percepción obtenida en tales casos.

Uno de los primeros problemas que surgen al comparar el ojo con una cámara fotográfica es el de que, con esta analogía, no se tiene en cuenta la cuestión de cómo están organizados los elementos que componen la imagen retiniana. Considérese el caso de tres puntos que no estén en línea recta. Sólo se apreciarán como los vértices de un triángulo si el sistema perceptual los agrupa en una unidad; en presencia de otros puntos, su triangularidad puede que sea menos evidente. Incluso una figura bien conocida compuesta de líneas puede no ser reconocida si el sistema perceptual asocia las líneas con otras líneas para formar diferentes agrupaciones. Y es lo que pasa al organizar un dibujo distinguiendo entre figura y fondo: una forma familiar puede no percibirse –aun cuando una imagen de esa forma esté presente en la retina– si se toma por figura alguna otra parte del campo visual.

Otro problema es el de la orientación. Si no se percibe bien dónde están la parte superior, la parte inferior y las partes laterales de una figura, la mera presencia de su imagen en la retina no garantiza el reconocimiento de la figura. La orientación de una figura en la tercera dimensión puede constituir también un factor de la percepción de la forma. Por ejemplo, un círculo mirado oblicuamente proyecta sobre la retina una imagen elíptica, mas no por ello dejará de parecer circular en las condiciones adecuadas.

Todos estos ejemplos indican que la existencia de una exacta imagen retiniana no es una explicación suficiente de la percepción de una forma. Se han de invocar otros factores. A decir verdad, la exacta imagen retiniana no es, en algunos casos, factor necesario. Se demuestra fácilmente que, con la presencia de sólo unas partes de una figura, otras partes pueden percibirse como si estuviesen también presentes. Considérese el simple trazo lineal de dos rectángulos, uno de los cuales parece superponerse y tapar al otro; la figura se puede ver como dos rectángulos completos, aunque la imagen de uno de ellos sea incompleta. En otros casos, ciertos perfiles (llamados perfiles subjetivos o ilusorios) se distinguen a pesar de que en la imagen retiniana no se dé la brusca transición de la luz a la oscuridad que suele considerarse como el estímulo necesario para la percepción de un perfil.

En los dos últimos ejemplos, algunos de los trazos de la figura están presentes en la imagen retiniana. En otros experimentos, se puede percibir una forma sin que haya ningún trazo de la misma en la imagen retiniana. Con ayuda de un visor estereoscópico puede presentarse a cada ojo una distribución aleatoria de elementos gráficos, de tal suerte que con un solo ojo no resulte visible ninguna figura. Sin embargo, si un grupo de los elementos de la muestra puesta ante un ojo se desplaza horizontalmente con respecto al mismo grupo de la muestra puesta ante el otro ojo, el sistema perceptual puede detectar la semejanza de los dos grupos e inferir que hay una forma frente al resto de la muestra, o detrás de él. Aquí, la forma percibida emerge, no de los trazos retinianos presentes en cada ojo, sino de los confines de una región que el sistema perceptual infiere que es una entidad que hay en la escena.

Otro ejemplo tiene una conexión más estrecha con mis recientes trabajos. Supóngase que un punto luminoso recorre un trayecto en una habitación oscura y se pide a un observador que lo siga con la mirada. La imagen del punto caerá siempre aproximadamente en el mismo sitio de la retina (o sea, en la fóvea), pero el observador percibirá la forma del trayecto. Esta impresión viene a ser muy afín a la percepción de una figura, por más que en la retina no haya habido ninguna imagen extensa correspondiente a la trayectoria seguida por el punto. Al efectuar Fred Halper y yo este sencillo experimento en nuestro laboratorio de Rutgers, nos encontramos con que los observadores se percataban tan bien de la forma de tal trayectoria como cuando mantenían quietos los ojos y la imagen del punto recorría la retina.

Cabría objetar que la percepción de una trayectoria no es lo mismo que la percepción de una forma. Después de todo, en este último caso se ve un contorno, y no precisamente un punto en movimiento. Sin embargo, cambiando un poco el procedimiento experimental, hay una manera de conseguir que una forma produzca una fuerte impresión. En vez de la habitación oscura y el punto en movimiento, preséntese al observador una estrecha rendija abierta

en una superficie opaca que se vaya moviendo a lo largo de un trazo ondulado [*véase la ilustración de esta misma página*]. Si el observador sigue atentamente el movimiento de la rendija, lo que aparece a través de ésta estimula sólo una estrecha franja de su retina, y la imagen de la porción visible de la figura va, por consiguiente, subiendo y bajando dentro de esta franja retiniana. También puede moverse en vaivén la figura por detrás de una rendija estacionaria; si los ojos del observador permanecen en este caso quietos, mirando fijamente la rendija, también estimula ésta tan sólo una estrecha franja retiniana. Ni por un instante, en ninguno de los dos casos, hay en la retina imagen alguna que se parezca a la forma de la figura, del trazo ondulado. Así y todo, en estas condiciones, comprobamos que las figuras se perciben sin mayor dificultad.

En el siglo XIX varios investigadores, entre ellos Joseph Plateau, Fritz Zöllner y Hermann von Helmholtz, emplearon un dispositivo similar para estudiar ciertos fenómenos visuales. Se refirieron a su procedimiento llamándolo presentación anortoscópica, presumiblemente porque era un método de estimulación visual distinto de los acostumbrados. El ingenio experimental, denominado anortoscopio, consistía en dos discos montados axialmente uno detrás del otro y que podían girar a diferentes velocidades y en sentidos opuestos. El observador veía las figuras trazadas en el disco más distante mirándolas a través de rendijas abiertas en el disco más próximo, con lo que percibía ciertas deformaciones ilusorias de las figuras.

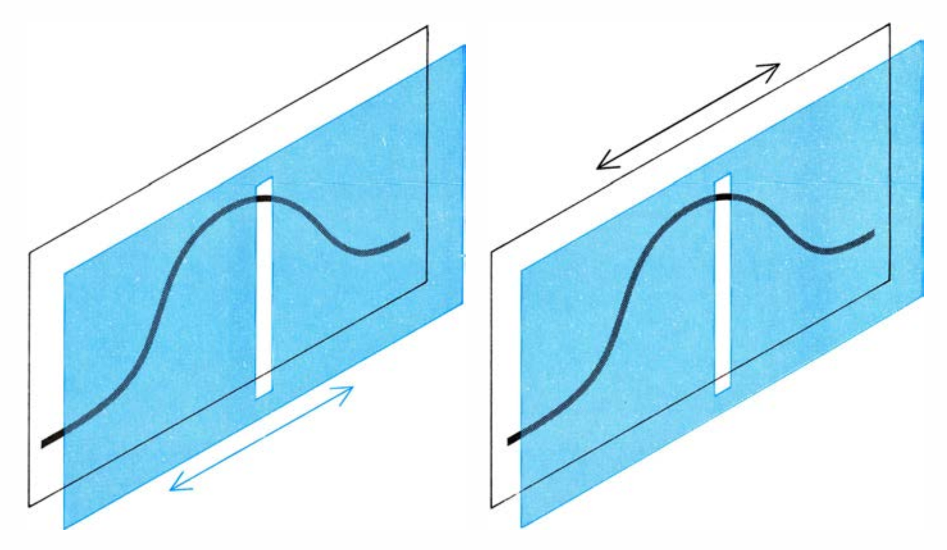
Este primer trabajo no recibió atención alguna, según parece, hasta 1965, año en que lo redescubrió T. E. Parks, de la Universidad de California en Davis. Tanto en el siglo XIX como después del redescubrimiento, se entabló una controversia sobre cómo interpretar el efecto. Helmholtz y otros pensaron que en la retina se imprimía siempre una imagen de la figura situada tras la rendija. A esta interpretación se la conoce desde entonces como hipótesis de la "impresión retiniana", por suponer que, sobre la retina, se va extendiendo cada vez una porción de la imagen. Si la rendija se mueve por delante de una figura estacionaria, la impresión retiniana exigirá que el observador mantenga quietos los ojos, sin seguir con la mirada el movimiento de la rendija. Si se mueve la figura por detrás de la rendija estacionaria, la impresión retiniana requerirá que el observador desplace la mirada. Helmholtz argüía que el observador no siempre podría advertir si sus ojos se mantenían quietos o se movían.

De ser correcta la hipótesis de la impresión retiniana, tendría poco de sorprendente, en tales condiciones, la percepción de la forma. En los dos casos arriba citados se daría entonces una extensa imagen retiniana de la figura, y la única diferencia entre la percepción en la situación visual anormal y en la normal consistiría en que, en la visión anortoscópica, la imagen se iría creando por segmentos, presentados, no simultánea, sino sucesivamente. Más aún, como las células sensoriales de la retina tienden a seguir descargándose durante un breve tiempo una vez terminada la estimulación (fenómeno conocido por el nombre de persistencia neuronal), la imagen presentada en segmentos sucesivos no diferiría gran cosa de una imagen formada toda ella al mismo tiempo, siempre y cuando las presentaciones sucesivas fuesen lo bastante rápidas. Sería, pues, de esperar que el observador percibiese la imagen anortoscópica entera como si se hubiese establecido normalmente.

Para evaluar como es debido la hipótesis de la impresión retiniana hay que saber con exactitud qué es lo que hacen los ojos del observador durante una presentación anortoscópica. De ahí que en nuestros experimentos solamos registrar en cintas de vídeo los ojos del observador durante la presentación. En una prueba, Joseph Di Vita y yo hicimos que los observadores mirasen las figuras de unas líneas onduladas de aproximadamente 14 centímetros de longitud a través de una rendija de unos tres milímetros de anchura desde una distancia de unos 75 centímetros. En cinco ciclos consecutivos se movía en movimiento de vaivén, ya la figura, ya la rendija. La velocidad de la figura o de la rendija era de un pase (o sea, medio ciclo) por segundo, velocidad inferior a la adoptada por otros investigadores de la percepción anortoscópica. Los registros de la actividad ocular indicaban en cada momento la dirección de la mirada con respecto al dispositivo. Al observador se le pidió que pulsara un conmutador solamente durante los períodos en que percibiese una figura tras la rendija. La señal se registraba junto con el registro del movimiento del ojo en la cinta de vídeo.

Cuando la rendija se mantenía estacionaria, los datos mostraban que, por lo general, los observadores mantenían también quietos los ojos, independientemente de que se les hubiese dicho o no que los mantuvieran así y de que se les indicase o no un punto de la rendija donde fijar la mirada. Naturalmente, los ojos no quedaban perfectamente quietos, pero sus pequeños y rápidos movimientos ocasionales hacia uno u otro lado (la llamada titilación del ojo) no eran suficientes para explicar la percepción de la forma. Los observadores no siempre distinguían una figura, pero en muchas ocasiones sí lo hacían, y sus ojos estaban más o menos estacionarios.

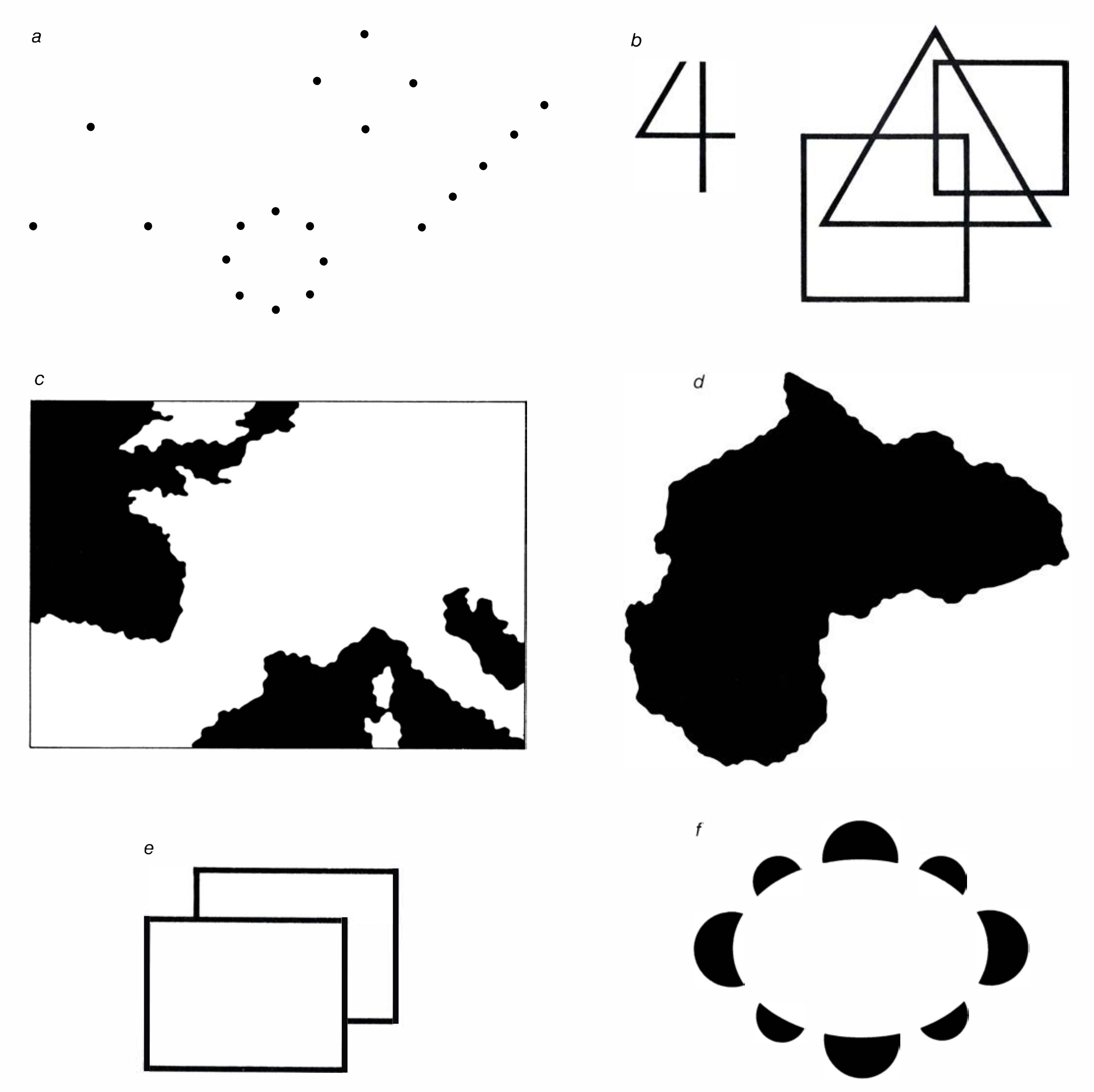
Estas averiguaciones fueron confirmadas recientemente por Robert Fendrich y Arien Mack, de la Nueva Escuela de Investigación Social. En su



PRESENTACION ANORTOSCOPICA, método anormal de estimulación visual, adoptado por el autor para someter a prueba la hipótesis de la imagen retiniana, que sostiene que en la retina se impresiona cada vez un área distinta. El procedimiento requiere mirar las secciones sucesivas de una figura por una rendija. Dos son las condiciones experimentales: o se mueve una rendija por delante de una figura estacionaria (*izquierda*), o bien se mueve una figura por detrás de una rendija estacionaria (*derecha*).

experimento, la imagen expuesta se estabilizaba de suerte que, aun cuando se movieran los ojos, no podría desplazarse por la retina la imagen de la rendija. Sin embargo, sus observadores percibieron también una figura que se movía tras la rendija.

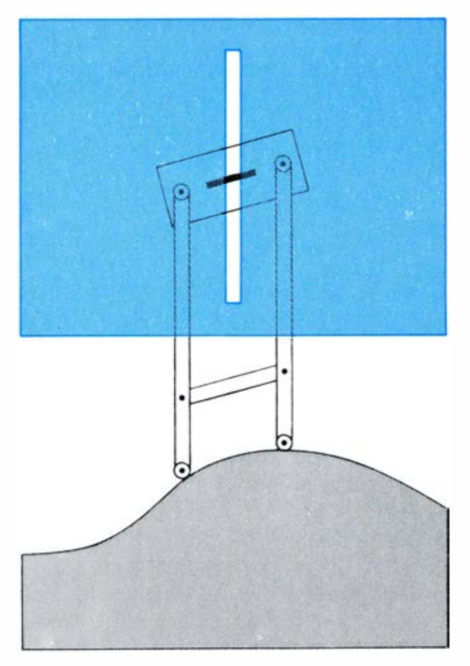
Cuando en nuestros experimentos se movía la rendija sobre una figura estacionaria, los observadores movían característicamente sus ojos siguiendo el recorrido de la rendija. Pero solía haber un momento en el que la rendija iba más deprisa que los ojos y mantenía su ventaja hasta llegar a la posición terminal, donde los ojos la alcanzaban con rapidez. Estaba claro, no obstante, que la imagen de la rendija no se desplazaba sobre la retina lo suficiente como para explicar la percepción de la forma; a decir verdad, la imagen retiniana ni siquiera se movía de continuo en el sentido idóneo. En esta situación, se percibía casi siempre una figura, y su tama-

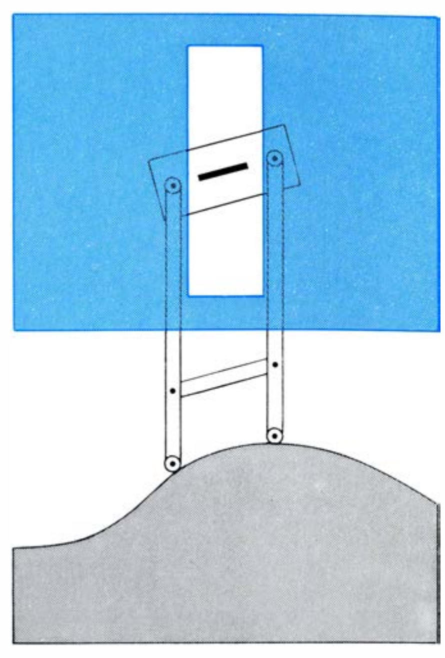


LIMITES DE LA FUNCION DE LA IMAGEN retiniana en la percepción visual de la forma. Se ilustra por medio de seis ejemplos. En *a*, los tres puntos de la izquierda sólo se aprecian como los vértices de un triángulo si el sistema perceptual los organiza como una unidad; en presencia de otros puntos (*derecha*), el triángulo no se percibe. En *b*, el número 4 (*izquierda*) no se distingue como tal si el sistema perceptual asocia sus líneas constitutivas con otras líneas en la formación de diferentes agrupaciones (*derecha*). En *c*, la familiar forma de Europa occidental no se percibirá si se toma por figura el fondo (en este caso las extensiones de agua que la rodean). Análogamente, en *d* no se reconoce el continente africano si no se cae en la cuenta de que lo que debería ser la parte superior de la figura (o sea, el norte) está aquí como parte izquierda. Estos cuatro ejemplos demuestran que la mera presencia de una determinada imagen retiniana es insuficiente para explicar lo percibido. En *e*, dos formas rectangulares, una superponiéndose y tapando en parte a la otra, se distinguen como dos rectángulos completos aunque uno de ellos esté, de hecho, incompleto. En *f* se percibe fácilmente un contorno subjetivo (o ilusorio), aunque falta la brusca transición de la luz a la oscuridad que suele considerarse como estímulo necesario para la percepción de contornos. Los dos últimos ejemplos ponen de manifiesto que no es menester que haya una imagen exacta o completa para explicar lo que se percibe. (Ilustración de Jerome Kuhl.)

ño se apreciaba más o menos correctamente.

Otros experimentos realizados en nuestro laboratorio y en otros centros, indicaban que la única manera de que el efecto anortoscópico pudiese resultar de la impresión retiniana sería si la imagen de la rendija se moviese en la retina muy rápidamente. Sólo a una velocidad próxima a las cinco pasadas por segundo, o sea, cinco veces mayor que la que nosotros solíamos emplear, podría explicarse la impresión de una figura entera por persistencia neuronal. Añádase que, en la condición de que se moviese la figura, la impresión retiniana tendría que basarse en los movimientos del ojo, y por tanto los ojos deberían moverse de un lado a otro con celeridad. Los observadores estudiados fueron incapaces de mover sus ojos a tal velocidad.





TECNICA DE SIMULACION ideada por el autor y por uno de sus colegas, Alan L. Gilchrist, para estudiar el fenómeno de la percepción anortoscópica. Este artificio se emplea para comprobar que, para que una figura se perciba en condiciones anortoscópicas, la porción visible de la figura ha de extenderse a través de toda la rendija. Un pequeño segmento de línea trazado en una pieza de plástico transparente se sujeta con unas varillas a un conjunto de rodillos que giran sobre una leva no visible por la rendija. La parte visible del segmento de línea se mueve arriba y abajo; simultáneamente, su inclinación va cambiando exactamente igual que lo haría la parte visible de una figura lineal auténtica. Cuando la rendija es más estrecha que la longitud del segmento, se percibe a menudo una figura (*izquierda*). Sin embargo, cuando la rendija es más ancha que largo el segmento, el observador no percibe nunca la figura (*derecha*).

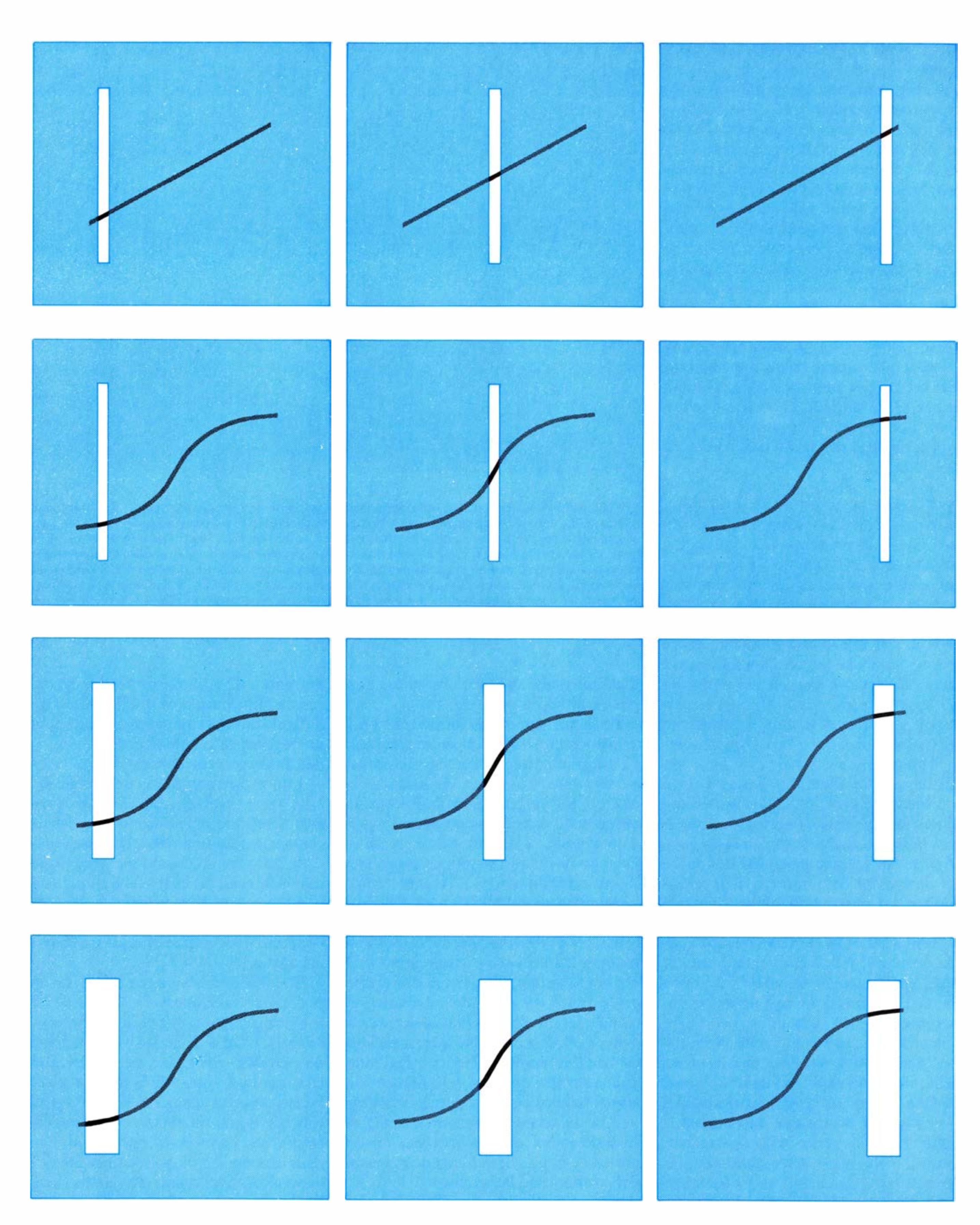
Basándonos en esta y en otras pruebas, concluimos que la impresión retiniana no constituía una explicación suficientemente general de la percepción anortoscópica. ¿Cuál era entonces la base de la percepción de la forma en los experimentos de esta especie? Considérese de nuevo el seguimiento de un punto luminoso en una habitación oscura. Aunque la imagen del punto no se mueve en la retina, el sistema perceptual recibe de hecho información acerca de la trayectoria del punto, porque han de moverse los ojos para irlo siguiendo. El sistema perceptual toma en cuenta la posición de los ojos cuando estima, a partir del estímulo retiniano, la situación del punto en el espacio. Así, la trayectoria percibida del punto se deriva del conjunto de las localizaciones del punto percibidas a medida que éste evoluciona.

Generalizando el razonamiento a toda percepción de forma, puede decirse que, cuando está presente en la retina la imagen extensa de una figura, el observador percibe cada punto de la figura como situado en un lugar determinado. Cabe, pues, suponer que la forma percibida es el resultado de una síntesis mental de las localizaciones percibidas de todos los puntos que constituyen la figura. Este proceso quizá se organice jerárquicamente, sintetizándose los puntos para formar líneas y disponiéndose éstas también en síntesis unas respecto a otras para formar las figuras. De ser así, tal vez fuera el conjunto de localizaciones percibidas lo que tuviese importancia al tratar de la imagen retiniana de una figura. En otras palabras, quizás el papel de los contornos retinianos de una figura sea fijar una serie de localizaciones que, en conjunto, den por resultado una forma.

Podemos aplicar el mismo razonamiento a la percepción anortoscópica. El sistema perceptual integra la información que, acerca de las localizaciones de las partes de la figura, se va obteniendo a medida que aparecen sucesivamente a través de la rendija. Cuando la rendija se mueve sobre una figura estacionaria, el elemento visible a un extremo de la figura es descrito, digamos, como "en lo más bajo de la rendija y hacia la izquierda", y sólo un momento después se distingue ya al elemento visible como "algo más arriba en la rendija y enfrente". La integración de las direcciones lleva a la percepción de una línea oblicua que va subiendo hacia la derecha. Por consiguiente, puede decirse que, tanto la percepción anortoscópica como la percepción normal de la forma requieren la integración de todas las localizaciones de las partes de una figura en una descripción del conjunto de las mismas.

Esta explicación resulta incompleta por varias razones. La primera es que la dirección respecto al observador se da directamente en el supuesto de que se mueva la rendija, pero no en el de que se mueva la figura. En este último caso la rendija permanece quieta y enfrente. ¿De dónde proviene, entonces, la información de que las partes de la figura sucesivamente reveladas están en lugares diferentes? Parece que lo que primero cabe suponer es que algo se está moviendo detrás de la rendija. Una vez concebida esta idea, la sucesión temporal de los estímulos puede convertirse en una configuración espacial. Después, lo mismo que cuando se movía la rendija, el sistema perceptual puede integrar la secuencia de los estímulos en una forma que se basa en el conjunto de localizaciones de las distintas partes.

Nuestro análisis sugiere otra razón por la que la hipótesis de la síntesis de las localizaciones no acaba de explicarlo todo. A menudo, los observadores no perciben una figura en la presentación anortoscópica. Este resultado debe de guardar alguna relación con la clase de figura empleada por nosotros, puesto que otros investigadores no han dado noticia de tales fracasos. Hemos escogido deliberadamente figuras dotadas de varios rasgos distintivos: son líneas únicas y que están trazadas de tal modo que sólo un elemento de la figura pueda verse cada vez a través de la rendija; son curvas bastante suaves, no trazos con bruscos cambios de dirección; son líneas continuas, no intermitentes o rotas; y no representan objetos familia-



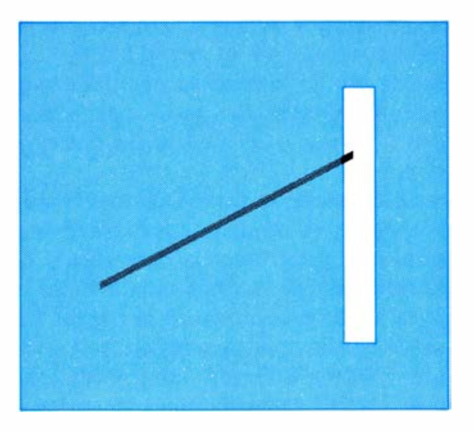
LA FUERTE TENDENCIA a percibir una figura lineal expuesta anortoscópicamente como un segmento de línea que se desplaza verticalmente y no como una línea continua que se revele en sentido horizontal se demuestra en cuatro secuencias de exhibición, representando cada una tres momentos sucesivos. Cuando una línea recta oblicua se presenta anortoscópicamente (*arriba*), nunca se percibe como una figura extensa en movimiento, presumiblemente porque no hay indicio de cambio en la identidad del segmento. Cuando una figura curvilínea se presenta en las mismas condiciones (*segunda desde arriba*), el segmento visible se acelera y decelera a medida que se va moviendo verticalmente en la rendija, pero esta información adicional no parece cambiar la percepción. Aunque se doble la anchura de la rendija, de modo que se detecte la inclinación y el cambio de inclinación del segmento visible (*tercera desde arriba*), la mayoría de los observadores todavía no perciben una figura. Sólo cuando la rendija se ensancha aún más, de modo que se detecte la curvatura y el cambio de la misma (*secuencia inferior*), se percibe una figura. Todas las presentaciones se observaron a unos 90 centímetros de distancia.

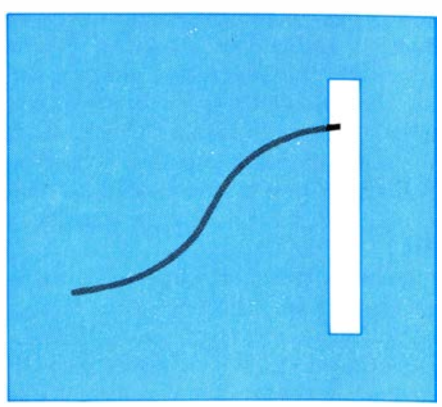
res. Otros han empleado cuadrados, círculos y elipses o ligeros apuntes de objetos muy conocidos, por ejemplo de animales.

Consideramos que nuestra especie de figura es ambigua, pues su exhibición anortoscópica podía representar lógicamente, ya un elemento moviéndose en sentido vertical en una rendija, ya una figura extensa que se entreviera a través de la rendija. Por ello, decidimos investigar qué factores cobraban importancia en la percepción de la figura, pues no siempre se percibía ésta. Nuestra hipótesis de trabajo fue que la percepción de la figura, en condiciones anortoscópicas, se basaba en un proceso inconsciente de resolución de un problema. Aun antes de que empiece cualquier actividad sintetizadora de localizaciones, el sistema perceptual ha de dar por supuesto que tras la rendija hay una figura extensa y que se está mostrando por secciones sucesivas.

Para aclarar esta hipótesis, considérese lo que se vería si la figura fuera luminosa y se observase en una habitación oscura, de modo que la rendija no pudiera apreciarse. En la situación de rendija estacionaria, todo lo que vería el observador sería un punto de luz moviéndose arriba y abajo. En la situación de rendija móvil, el observador vería un punto en movimiento que recorrería una trayectoria. En ninguno de los dos casos se describiría el percepto (la entidad percibida) como una figura extensa. Parece, pues, que el observador tiene que caer en la cuenta de que la momentánea visión se efectúa a través de una rendija y de que el resto de la figura queda oculta por una superficie opaca. A diferencia de la resolución de un problema, como acto de pensamiento, la resolución de un problema de percepción requiere el soporte del estímulo. Para percibir anortoscópicamente una figura no basta con saber que se está mirando a través de una rendija; hay que ver la rendija.

Bajo ciertas condiciones, el observador puede incluso "inventarse" una rendija, cuando no hay ninguna visible. Fendrich y Mack han constatado que, si se mueve un círculo tras una rendija invisible y se hace la sugestión apropiada, observadores que primero apreciaban las partes visibles del círculo en movimiento vertical perciben después una figura que se mueve horizontalmente por detrás de una rendija ilusoria. Este experimento proporciona un buen ejemplo de cómo el sistema perceptual tiende a racionalizar una solución adoptada por la facultad de resolver problemas, si el estímulo no la apoya plenamente.





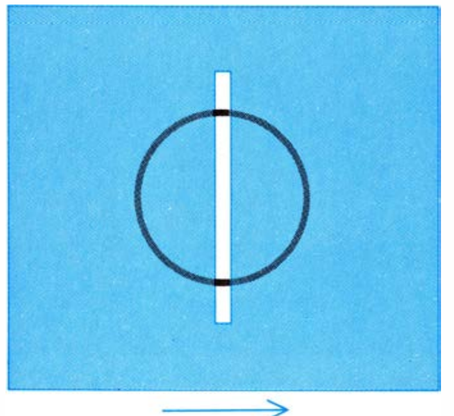
APARICION DE LOS PUNTOS EXTREMOS de una línea en la rendija de un dispositivo anortoscópico. Aunque este factor puede ayudar a que se perciba una forma, no parece ser suficiente para inclinar la balanza a favor de la percepción de la figura. Una línea recta sesgada tiende a percibirse como un segmento de línea que se desplaza verticalmente aun cuando sus puntos extremos sean visibles (*izquierda*). Pero sí es más probable que una línea curva con su extremidad visible se aprecie como figura (*derecha*).

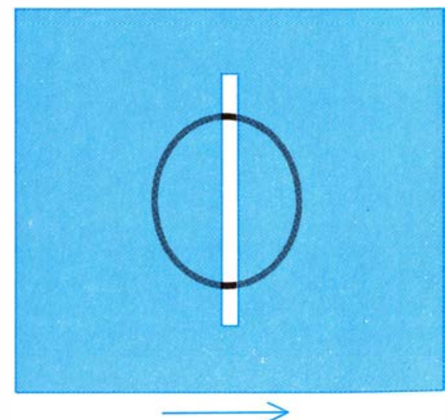
No sólo ha de verse la rendija; la porción visible de la figura ha de extenderse también por la totalidad de la rendija. Para aclarar este punto, Alan. L. Gilchrist y yo ideamos un procedimiento simulador de la presentación anortoscópica. En vez de figura, pusimos tras la rendija sólo un breve segmento lineal dibujado en un plástico transparente. El plástico estaba unido mediante unas varillas a un juego de rodillos que giraban sobre una leva no visible a través de la rendija; la leva tenía la forma de la figura que se trataba de simular [*véase la ilustración de la página 89*].

En ciertas situaciones, la simulación efectuada con tal instrumento hace que se perciba la figura. En nuestro caso, sin embargo, ensanchamos la rendija de suerte que el segmento lineal visible no ocupara toda su anchura, sino que quedase aislado en su interior. Aunque el segmento se movía verticalmente y su inclinación cambiaba lo mismo que cuando la rendija era más estrecha, en ningún momento se percibió figura alguna. El estímulo violaba manifiestamente la lógica de la solución: el que se viera aparecer a través de la rendija una figura continua.

Considérese luego la clase de figura que se mueve tras la rendija estacionaria. Si la figura es una línea recta sesgada, cuyos extremos no se ven a través de la rendija (siendo la anchura de ésta de alrededor de un milímetro y medio, y si se la mira a la distancia de unos 90 centímetros), el estímulo consiste en un segmento de línea que se va desplazando en sentido vertical. Este estímulo podría resultar de uno cualquiera de los siguientes hechos físicos: de que un segmento se moviera arriba y abajo con velocidad uniforme, o de que una línea oblicua se moviera en vaivén con velocidad uniforme.

Bajo tales condiciones de ambigüedad, Hans Wallach, del Swarthmore College, demostró hace ya mucho tiempo que la forma de la abertura influye de un modo destacado en lo que se percibe. Hay una fuerte tendencia a perci-





EFECTO DE COMPRESION observado a menudo en despliegues anortoscópicos en los que la figura se mueve tras una rendija estacionaria. La figura percibida, en este caso es un círculo (*izquierda*), está típicamente comprimida en la longitud de su eje de movimiento, formando lo que parece ser una elipse.

bir la línea como moviéndose ésta en sentido paralelo al eje mayor de la abertura, probablemente porque con una rendija estrecha no hay información que sugiera cambio alguno en la identidad del segmento. Así, hemos hallado que, bajo tales condiciones, nunca se percibe una figura extensa y móvil.

Supongamos que la figura sea una línea curva y que las condiciones de presentación sean, por lo demás, las mismas. Entonces, el único factor nuevo es que el segmento visible se acelera y se desacelera al moverse verticalmente. Ann Corrigan y yo hemos comprobado que esta información no altera el resultado. Todos los observadores perciben un segmento que se mueve verticalmente, no una figura.

Supóngase a continuación que se aumenta, hasta ser de unos tres milímetros, la anchura de la rendija. En estas condiciones, vemos que la inclinación del segmento visible y el cambio de inclinación pueden detectarse a través de la rendija. Sin embargo, aun con esta información y con las correlativas aceleración y deceleración en vertical del segmento, la mayoría de los observadores no perciben todavía una figura. Cabría concluir de ello que la tendencia, advertida por Wallach, a que un segmento mantenga su identidad es tan fuerte que la seguirá conservando por más que el segmento parezca inclinarse a un lado y a otro en sus subidas y bajadas.

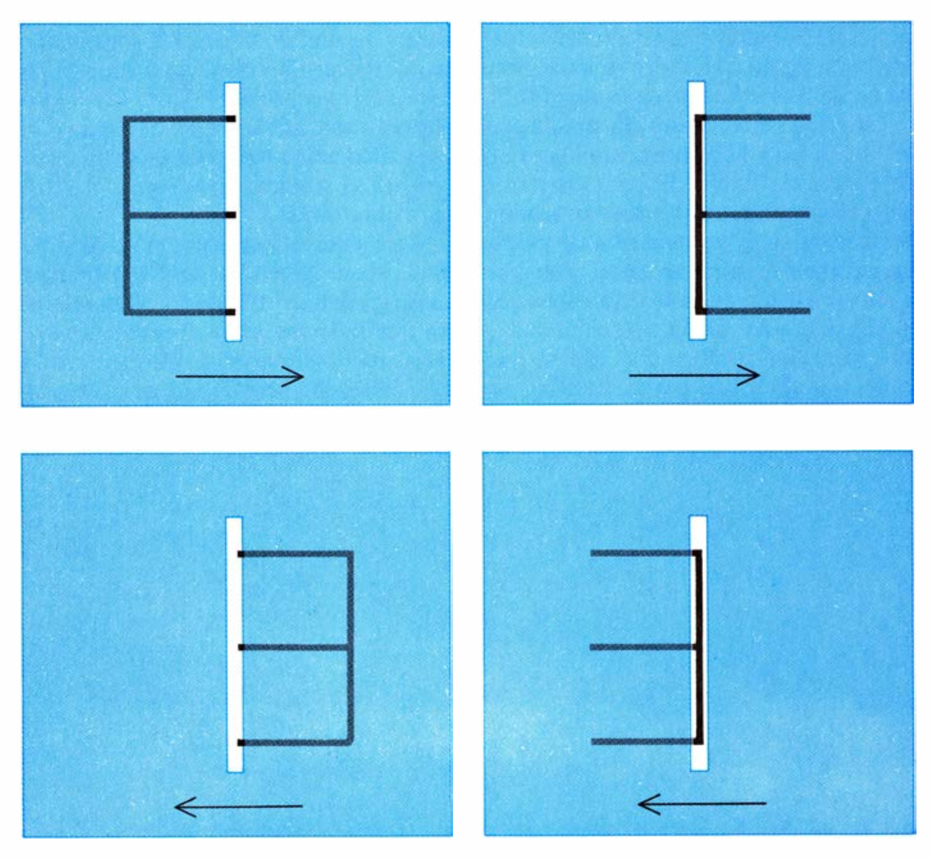
Sólo cuando se ensancha todavía más la rendija, de modo que a su través resultan ya detectables la curvatura y el cambio de ésta, se inclina la balanza a favor de la percepción de la figura. La razón de ello no está todavía clara, aunque vienen a la mente varias posibilidades. Cabría especular que la aceleración del segmento de línea ascendente justo cuando su pendiente se va acercando a la vertical, a la vez que su curvatura cambia proporcionalmente, representaría una coincidencia demasiado grande si por detrás de la rendija no se extendiese ninguna línea curva. Una solución de figura móvil explica por qué estos rasgos varían en conjunto, y el sistema perceptual tiende a buscar tales soluciones y a rechazar perceptos que postulen variaciones meramente casuales.

Podría también decirse que la noción de segmento lineal que se mueve verticalmente al propio tiempo que se ladea y que cambia de curvatura es una percepción muy compleja. Una figura móvil es una percepción más simple, y se ha argüido que el sistema perceptual prefiere la descripción más simple que concuerde con el estímulo. Cabe asimismo conjeturar que la combinación de cambios, incluidos los de la curvatura, resulta reconocible para el sistema perceptual, de acuerdo con la experiencia, como la combinación producida por un objeto al pasar tras una estrecha abertura.

Se dan en el proceso otros tipos de información que pueden afectar al resultado. Dejar que se vean a través de la rendija los extremos de la figura puede llevar a la percepción de una forma. Desde luego, cuando se aprecian los extremos, hay un momento de información inequívoca en el sentido de que algo se mueve cruzando la ranura, no a lo largo de ella. Podría interpretarse esta información como un indicio que sugiere la hipótesis de la figura, pero, en sí misma, la información no parece ser adecuada; una línea recta oblicua tiende todavía a verse como un segmento que se mueve en sentido vertical, excepto durante el breve tiempo en que están a la vista sus dos extremos. Cuando la información de fin-de-línea se acopla, empero, con la información de cambio-de-sesgo, la balanza se inclina a menudo hacia la percepción de una figura. Quizá la razón sea que la percepción de un cambio de sesgo tienda a corroborar la hipótesis sugerida por la visión de los extremos de la línea. El tipo de información inequívoca suministrada por los puntos terminales tal vez haga totalmente presente una figura, si ésta tiene discontinuidades o cambios bruscos en la dirección de sus contornos.

Hay otro factor cuya importancia hemos apreciado. Si el observador sabe que la demostración puede representar una figura tras una rendija, es más probable que la perciba. Ordinariamente, el mero hecho de saber algo (por ejemplo, que una ilusión es una ilusión) no afecta a lo que se percibe; pero en este caso sí. Aun con conocimiento previo, las condiciones han de ser favorables para que se dé percepción de forma. Ese conocimiento no influye si la inclinación del segmento lineal no cambia o si la figura es luminosa y se la mira en la oscuridad sin que se aprecie la rendija.

Nuestra investigación sobre las características del proceso que rige el resultado perceptual gira en torno a la cuestión de si el efecto anortoscópico es genuinamente perceptual, cuando se descarta la imagen retiniana. Varios estudiosos han pretendido que, cuando



LA FIGURA PERCIBIDA SE INVIERTE si un observador estima erróneamente el sentido de su movimiento tras una rendija estacionaria. En este dispositivo experimental, la letra *E* aparece si el sentido del movimiento de la figura es percibido correctamente hacia la derecha (*arriba*), pero se ve la figura invertida Ǝ si se interpreta erróneamente que el sentido del movimiento es hacia la izquierda (*abajo*).

en cada uno de los sucesivos momentos se ve sólo parte de la figura, la impresión que se tiene de la figura total es más una consecuencia del conocimiento de su presencia que el resultado de una verdadera percepción de la forma. Sólo si se ve de una vez la figura entera –dicen– deberá describirse la experiencia como percepción de forma.

Opino que hay aquí una confusión que se deriva de un prejuicio teórico. Si se considera que la percepción de la forma presupone una imagen retiniana extensa, cuyas partes están todas presentes simultáneamente, y si descartamos la imagen retiniana, entonces el efecto anortoscópico no puede ser perceptual. Pero no hay razón alguna que obligue a equiparar la percepción de la forma a la simultaneidad. Oír una melodía o pronunciar una frase es ciertamente algo perceptual, aunque lo percibido se extienda en el tiempo. Sólo si la figura o la rendija se movieran muy lentamente podría decirse con propiedad que no se llega a percibir la figura, aunque se sepa que está ahí. Análogamente, podría bloquearse la percepción de una melodía separando sus tonos con lapsos temporales muy largos.

Lo cierto es que quienes han experimentado en el laboratorio el efecto anortoscópico están seguros de que es éste un fenómeno perceptual. Una razón que abona su certeza son, indudablemente, las claras diferencias que se han observado bajo diferentes condiciones. De ahí que los fallos cobren mucha importancia. Los observadores pueden saber que se está presentando una figura, pero nunca la perciben a no ser que las condiciones lo permitan. En otras palabras, hay casos en los que es apropiado decir que el observador sabe que se está presentando una figura, pero no la percibe; estos casos, por contraste, ponen bien en claro que los otros casos son perceptuales. Más aún, en muchas ocasiones la percepción va pasando de una posibilidad a otra durante la presentación; un cambio así, bajo condiciones ambiguas, es una de las señales distintivas del proceso perceptual.

Hay una peculiaridad en la presentación anortoscópica que no he mencionado todavía. Cuando la rendija se mantiene quieta y se mueve la figura, la figura percibida suele sufrir deformaciones; lo típico es que se comprima a lo largo del eje de su movimiento. Así, un círculo puede aparecer como una elipse con su eje mayor vertical. ¿A qué se debe este efecto? Los defensores de la hipótesis de la imagen retiniana han sugerido que tal distorsión resulta de no poder mover el observador sus ojos en perfecta sincronía con la figura. Si el movimiento del ojo se retrasa con respecto a la velocidad del movimiento de la figura, o si no es tan amplio como éste, la imagen impresa en la retina se condensará, lo que explicaría muy bien la deformación.

Un experimento realizado en 1967 por Stuart M. Anstis y Janette Atkinson, de la Universidad de Bristol, parece corroborar tal hipótesis. Introducían un blanco móvil (un punto luminoso) que el observador debía seguir con la mirada. Variando la velocidad de este punto con respecto a la velocidad de la figura, Anstis y Atkinson lograban "imprimir" una imagen retiniana de la figura cuya forma iba cambiando. Por ejemplo, si la figura en movimiento era un círculo y el blanco que se seguía se movía a media velocidad, la imagen formada en la retina solía ser una elipse cuyo eje vertical era doble que el horizontal. Si la figura se movía hacia la derecha mientras el blanco en seguimiento se movía hacia la izquierda, la imagen creada solía ser el reverso de la producida por la percepción ordinaria. Los observadores informaban que veían figuras cuyas formas correspondían con precisión a la imagen retiniana establecida.

Este hallazgo nos intrigó a mis colegas y a mí durante algún tiempo, pues teníamos buenas razones para dudar de que la hipótesis de la imagen retiniana explicase el efecto anortoscópico o esas distorsiones perceptuales. Por fin se nos ocurrió otra explicación que parece cuadrar mejor con los hechos. Para entenderla conviene considerar primero con más detenimiento la percepción del movimiento figural cuando la rendija permanece estacionaria y la figura se mueve. Según la hipótesis de la imagen retiniana, lo que debería verse es la figura entera dentro de una amplia abertura rectangular; el rectángulo representa la imagen de la rendija, que también debería pintarse en la retina. Pero no es esto lo que se ve, sino una figura que se mueve tras la estrecha rendija.

La pregunta es: ¿Qué determina el que se perciba la velocidad a que se mueve la figura? Dada la estrechez de la rendija, difícilmente puede transmitirse información exacta acerca de la velocidad de la figura por el breve paso, a través de la rendija, de algunos elementos de contorno distinguible. Más aún, las figuras empleadas por nosotros no se componen de elementos distinguibles, si se exceptúan sus extremos. Así que la velocidad de la figura se representa, a lo sumo, ambiguamente. El que se perciba la longitud de la figura depende por completo de que se perciba su velocidad, al menos según la interpretación de la percepción anortoscópica como resolución de un problema. Puesto que el resultado más común, cuando no hay ningún blanco al que seguir, es que se percibe una compresión, concluimos que se subestimaba la velocidad. Ignoramos a qué se debe, pero es importante tener en cuenta que tampoco hay razón ninguna para esperar una percepción exacta de la velocidad.

Cuando se introduce un blanco de seguimiento, el sistema perceptual parece suponer que la figura se está moviendo a la velocidad de ese blanco. Puede formularse la hipótesis general de que el movimiento del ojo es, en condiciones anortoscópicas, un indicio del movimiento figural. Dado este indicio, la velocidad percibida de la figura se dobla cuando el blanco se mueve a doble velocidad que la figura, y se divide por dos cuando el blanco se mueve a la mitad de velocidad que la figura. Como la longitud aparente de la figura depende de cuán lejos veamos moverse ésta por detrás de la rendija durante el intervalo entre la aparición de un extremo y la aparición del otro, se explica la distorsión observada por Anstis y Atkinson. Resulta de una construcción mental de la longitud, deducida de la velocidad aparente de la figura, y no directamente de una imagen retiniana deformada.

Para aportar pruebas a nuestra interpretación, Di Vita, Halper, Deborah Wheeler y yo efectuamos un experimento en el que los observadores veían una figura curvilínea que se movía a cierta velocidad por detrás de una rendija y, a la vez, seguían con la mirada un determinado punto móvil. De hecho, repetimos el experimento de Anstis y Atkinson, pero con nuestro tipo de figura. Y había también una importante añadidura al procedimiento: el observador no sólo indicaba la longitud de la figura percibida (mediante la adaptación de un dispositivo sombreador que variaba la longitud de una réplica de la figura manteniendo su altura constante), sino que nos iba diciendo también si la velocidad de la figura parecía ser igual, menor o mayor que la del punto móvil. Se hicieron observaciones a distintas velocidades del punto: a igual velocidad que la de la figura, a la mitad de velocidad, al doble y a la misma velocidad pero en sentido opuesto al de la figura. Hubo también un ensayo de mi-

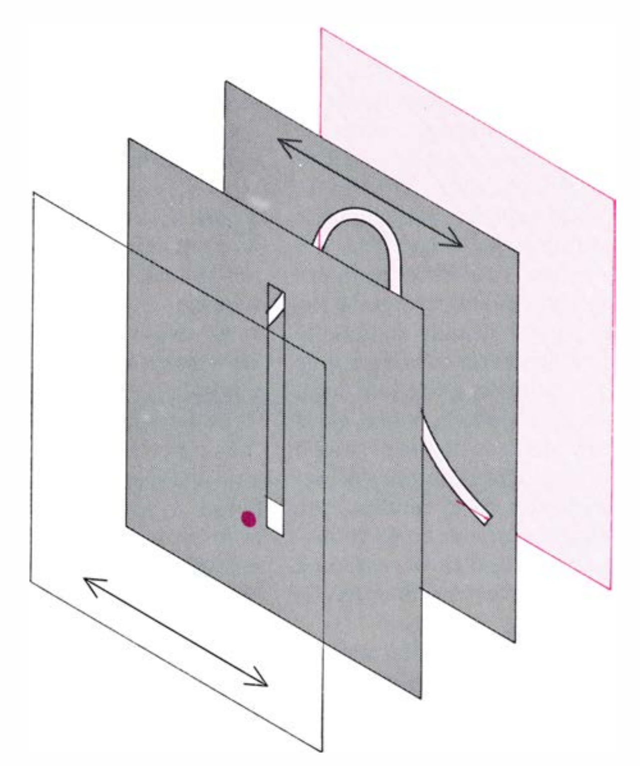
rada libre, sin presencia de ningún blanco móvil.

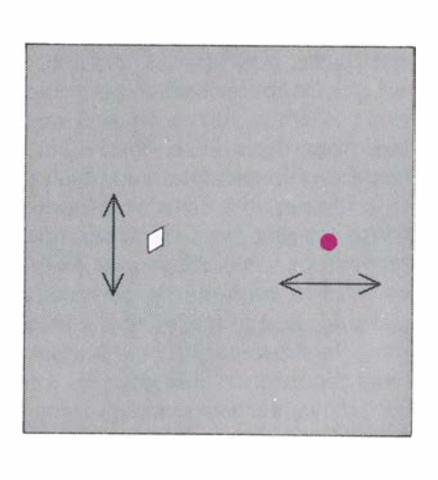
Siempre que tuvieron que seguir con la mirada algún blanco, los observadores percibieron una figura, mientras que en la condición de mirada libre no la percibió ninguno. En este experimento hicimos deliberadamente que la rendija fuese muy estrecha (de un milímetro y medio), lo que, según sabíamos, suele eliminar el efecto anortoscópico bajo la más usual condición de ausencia de blanco al que seguir.

Lo primero que hay que anotar en cuanto a los resultados es que el seguimiento visual de un blanco móvil es un factor importante en la percepción anortoscópica de la figura. Creemos que la principal razón de ello estriba en que el movimiento de los ojos proporciona un indicio efectivo de que, tras la rendija, se está moviendo, a un lado y a otro, una figura. Más aún, tal indicio especifica la velocidad de la figura, mientras que sin él la velocidad queda indeterminada. Sería difícil llegar a percibir una figura de una longitud definida siendo indeterminada su velocidad. Virtualmente todos los observadores estimaron que la velocidad de la figura era la misma, o casi la misma, que la del punto móvil, independientemente de cuál fuera la velocidad real de uno y otra. En cuanto a la longitud percibida de la figura, si tomamos como 1 la longitud real, la longitud percibida era de 0,77 cuando las dos velocidades eran iguales, 1,65 cuando la velocidad del punto móvil doblaba la de la figura y 0,50 cuando la velocidad del blanco era la mitad de la de la figura.

Cuando la figura se movía en un sentido y el punto en sentido opuesto, y a la misma velocidad, todos los observadores percibían la figura como la imagen especular de su verdadera forma. Para esta prueba empleamos una figura asimétrica y pedimos a cada observador que trazase un ligero esbozo de lo que percibía. El resultado fue el que se produciría si en la retina se imprimiese la imagen invertida, pero también era el que se produciría a partir de nuestra hipótesis del movimiento del ojo. No sólo resultaba ambigua la velocidad de la figura tras la rendija; también lo fue el sentido del desplazamiento. Por ejemplo, si la figura era la letra *E* y se movía hacia la derecha, los extremos de sus tres trazos horizontales se revelarían en primer lugar. Si la figura parecía moverse hacia la izquierda, pero los extremos de sus trazos aún aparecían primero e iban seguidos por el trazo vertical de la *E*, la figura tendría que ser una figura invertida: Ǝ. Que el sentido del movimiento se percibía erróneamente de este modo lo confirmó la respuesta incorrecta de todos aquellos observadores que percibían invertida la figura cuando se les preguntaba por el sentido del movimiento.

En este último experimento, el hecho de ir siguiendo con la mirada un blanco móvil producía el efecto de extender la imagen de la figura sobre la retina. Aunque hemos rechazado la hipótesis de que la imagen así impresa fuera la causa directa de la percepción anortoscópica de la forma y de las distorsiones observadas, hay un aspecto de los resultados que sugiere que la imagen extensa desempeña un papel importante. La figura percibida es más clara y más similar a la percepción ordinaria de la forma, bajo estas condiciones, que cuando los ojos se mantienen quietos al mirar la rendija estacionaria. Hemos dado con un resultado parecido cuando, en la situación de rendija en movimiento, se introduce un punto estacionario y el observador fija en él la mirada. En este caso se extiende también la imagen sobre la retina, mientras





SUPERFICIES empleadas por el autor y sus colegas en uno de sus experimentos de control sobre la percepción anortoscópica. La primera superficie, frente al observador, es una lámina de plástico transparente en la que va montando un blanco luminoso seguible (*punto de color*); esta lámina efectúa un movimiento de vaivén. La superficie siguiente es una cartulina opaca, con una rendija vertical practicada en su centro y que se mantiene estacionaria. La tercera superficie es otra cartulina opaca en la que se ha recortado la figura curvilínea que sirve de ensayo; esta superficie se mueve también en vaivén. La superficie de fondo es una lámina de plástico traslúcido que difumina la luz de la batería de lámparas puesta tras ella. El panel de la derecha muestra lo que ve el observador cuando se realiza el experimento en una habitación oscura: un blanco de seguimiento luminoso que se mueve horizontalmente y un segmento de línea que parece moverse en vertical (o casi en vertical). En estas condiciones no se ve una figura en movimiento, porque no se percibe la rendija. Cuando la figura se dibuja en la tercera superficie y las luces están encendidas, se ve siempre una figura que se mueve.

que cuando se sigue con la mirada la rendija en movimiento no hay tal imagen extensa. En ambos casos, cuando se presenta dicha imagen, es raro que el observador deje de percibir una figura aun cuando las condiciones sean, por lo demás, poco favorables para el efecto anortoscópico.

Al aumento de percepción anortoscópica atribuible a una imagen retiniana extensa lo llamamos nosotros efecto de facilitación. Aparentemente, al sistema perceptual le resulta más fácil integrar las porciones sucesivas de la figura en una forma total cuando las porciones se extienden por la retina que cuando caen todas sobre una banda vertical. Aunque la imagen extensa puede, pues, facilitar el efecto anortoscópico, no lo causa directamente. Las razones para llegar a esta conclusión quedan dadas más arriba, y la conclusión ha sido, además, corroborada por ulteriores experimentos de control.

En uno de estos experimentos, la figura era una línea recta horizontal, y el observador tenía que seguir con la mirada un blanco que se movía a igual velocidad que la figura. Aunque sobre la retina se extendió la imagen de una figura, ningún observador logró percibirla. En vez de ello, el segmento visible de la línea parecía formar parte de la rendija y no se producía ninguna impresión de que algo se moviese (salvo el punto que servía de blanco). Lo único que podemos sacar en conclusión es que sin desplazamiento vertical del trazo no hay razón alguna para que el sistema perceptual infiera que tras la rendija se está moviendo una figura extensa; de ahí que la imagen extensa, aunque presente, no se integre en una forma mentalmente construida.

En otro experimento control, consistía la figura en una plantilla en la que se había recortado una línea curva cuyo estarcido se iluminaba por detrás. Se dispuso también como blanco de seguimiento un punto luminoso, y el conjunto debía observarse en un ámbito, por lo demás, oscuro. Cuando el observador seguía con la mirada el blanco móvil, se extendía en su retina una imagen de la línea brillante. Sin embargo, no se percibía ninguna figura. Más bien, el segmento visible de la figura, que era en esencia un pequeño punto de luz, parecía moverse arriba y abajo, aunque su trayectoria parecía ligeramente desviada de la vertical.

Como el seguimiento tiene por efecto extender sobre la retina la imagen del fragmento visible, que se deje de obtener percepción de figura en este caso muestra muy a las claras que el efecto anortoscópico depende de que se perciba la rendija como una abertura en una superficie opaca. Sólo entonces se tiene el apoyo necesario para construir mentalmente una figura extensa.

En cualquier caso, parecen ser dos los factores que se combinan para que tenga éxito la percepción anortoscópica de la forma durante el seguimiento de un blanco móvil. El movimiento del ojo sirve de indicio de que una figura se mueve en ángulo recto con respecto a la rendija e imparte a la figura una velocidad no ambigua. La formación de una imagen extensa en la retina facilita la integración de la figura. Tal vez sea ese auxilio lo que ha inducido a algunos investigadores a declararse partidarios de la hipótesis de la imagen retiniana.

En conclusión, he sostenido que la percepción de la forma puede entenderse como un proceso integrador de información sobre la localización de las partes de una figura (siempre que las partes se organicen en una unidad y se interpreten como figura y no como fondo). Los trazos físicos o sus representaciones en la imagen retiniana no son necesarios siempre que algún tipo de información indique dónde están los límites de la figura. Por consiguiente, la presentación anortoscópica de una figura por porciones sucesivas de la misma puede producir la percepción de su forma aunque no aparezca en la retina, en algún momento dado, ninguna imagen extensa de la figura entera.

Sin embargo, puesto que la exhibición anortoscópica es ambigua y no representa necesariamente una figura encajada, el hecho de que se perciba una forma supone un proceso de resolución de un problema. Es indispensable para ello la presencia de una región que se interprete como una estrecha abertura en una superficie opaca que la rodee, y el estímulo ha de tener, además, algunas otras propiedades para que la solución preferida sea la percepción de una figura. Si una figura se mueve por detrás de una rendija estacionaria, el sistema perceptual ha de inferir también la velocidad y el sentido del desplazamiento de la figura para reconstruir su longitud y su forma.

Si esta interpretación de los eventos que se siguen al mirar un despliegue anortoscópico es correcta, la percepción de la forma es un proceso mucho más próximo al nivel cognitivo que lo que hasta ahora se venía reconociendo. No se puede explicar como un resultado directo de la transformación fisiológica de los trazos o contornos que estimulan la retina.

Bioquímica de la resistencia a la malaria

Los genes para dos enfermedades letales, la anemia falciforme y la talasemia, están favorecidos por la evolución, ya que protegen contra la malaria. Podemos estudiar ya en el laboratorio los mecanismos subyacentes a esta protección

Milton J. Friedman y William Trager

La evolución es el resultado de la selección natural, que actúa sobre una gama de diversidad genética que surge de la mutación y la recombinación de los genes. La frecuencia de los genes variantes que confieren alguna ventaja selectiva tiende a aumentar, mientras que los variantes deletéreos tienden a ser eliminados. En poblaciones humanas existen muy pocos ejemplos claros de selección a favor o en contra de genes específicos, como respuesta a fuerzas específicas. Los mejores ejemplos son las enfermedades hereditarias. La selección actúa contra los genes que causan tales enfermedades, y lo hace con mayor fuerza contra las condiciones más severas.

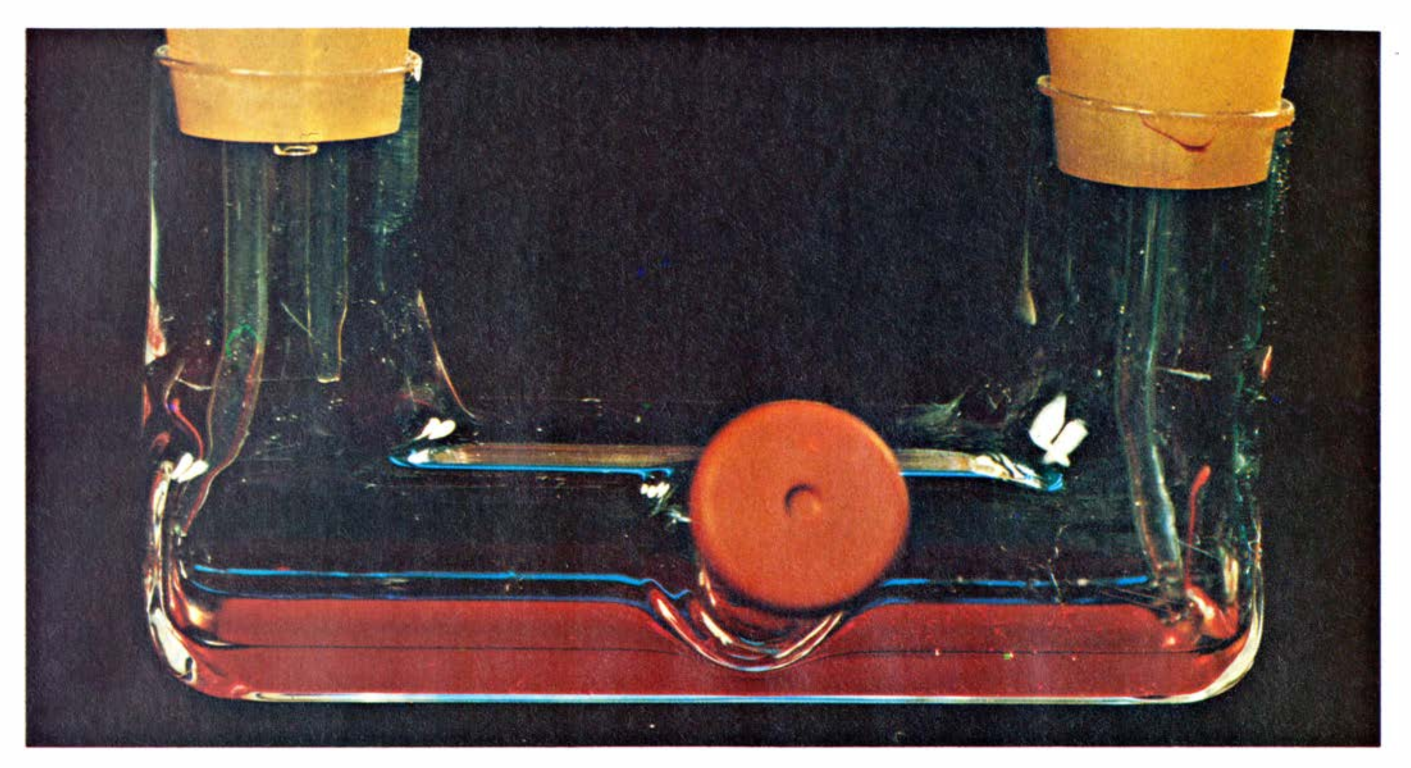
Siendo así, las enfermedades genéticas letales deberían ser muy raras. Sin embargo, ciertas anomalías hereditarias de los glóbulos rojos de la sangre, especialmente la anemia falciforme y la talasemia, se observan en algunas poblaciones con una frecuencia sorprendentemente elevada. ¿Vale ello de argumento en contra de la selección natural? Muy al contrario, los genes para la célula en forma de hoz y para la talasemia demuestran la fuerza de la selección en la evolución. Las mismas variantes génicas que causan la enfermedad letal en las células sanguíneas de individuos homozigóticos (que heredan dos de los genes anormales, uno de cada progenitor) protegen a los individuos heterozigóticos (que heredan un gen normal y otro anormal) frente a los efectos letales de la malaria. Esta protección mantiene la alta frecuencia de estos genes, de otro modo deletéreos.

La potencia de la malaria como fuerza selectiva deriva del poderoso efecto de la enfermedad parasitaria sobre la salud y sobre la capacidad reproductora de las poblaciones humanas. La malaria ha sido una de las principales causas de muerte a lo largo de la historia. Es hoy endémica en Africa. No barre las poblaciones como una epidemia; antes bien, constituye una calamidad constante que contribuye a la mortalidad infantil precoz, con tasas de hasta el 50 por ciento. Mata directamente a un 10 por ciento de sus víctimas e interviene en la muerte de otras disminuyendo la capacidad de su sistema inmunitario para luchar contra otras infecciones. Por su culpa, un número significativo de niños no alcanza la edad reproductora. Cualquier mutación genética que proporcione resistencia a la malaria debe tener, por tanto, una elevada ventaja selectiva.

Fue la coincidencia de la distribución geográfica de la enfermedad conocida por anemia falciforme con la distribución de la malaria lo que despertó la sospecha de que el gen de la anemia falciforme pudiera conferir tal resistencia. Aunque resultaba difícil obtener pruebas clínicas, en 1954 Anthony C. Allison, de la Universidad de Oxford, demostraba que los niños heterozigóticos para el gen de la anemia falciforme presentaban casos mucho menos graves de la forma más letal de malaria que los niños que no portaban el gen. Debido a que el parásito que causaba la malaria no podía mantenerse en cultivo de laboratorio, no se logró demostrar la resistencia a nivel celular, ni hubo manera de establecer su mecanismo bioquímico. Recientemente, hemos explotado un sistema de cultivo que acaba de desarrollarse, con el fin de ver cómo el gen de la anemia falciforme y algunos otros genes variantes que alteran las funciones de los glóbulos rojos confieren resistencia a la malaria.

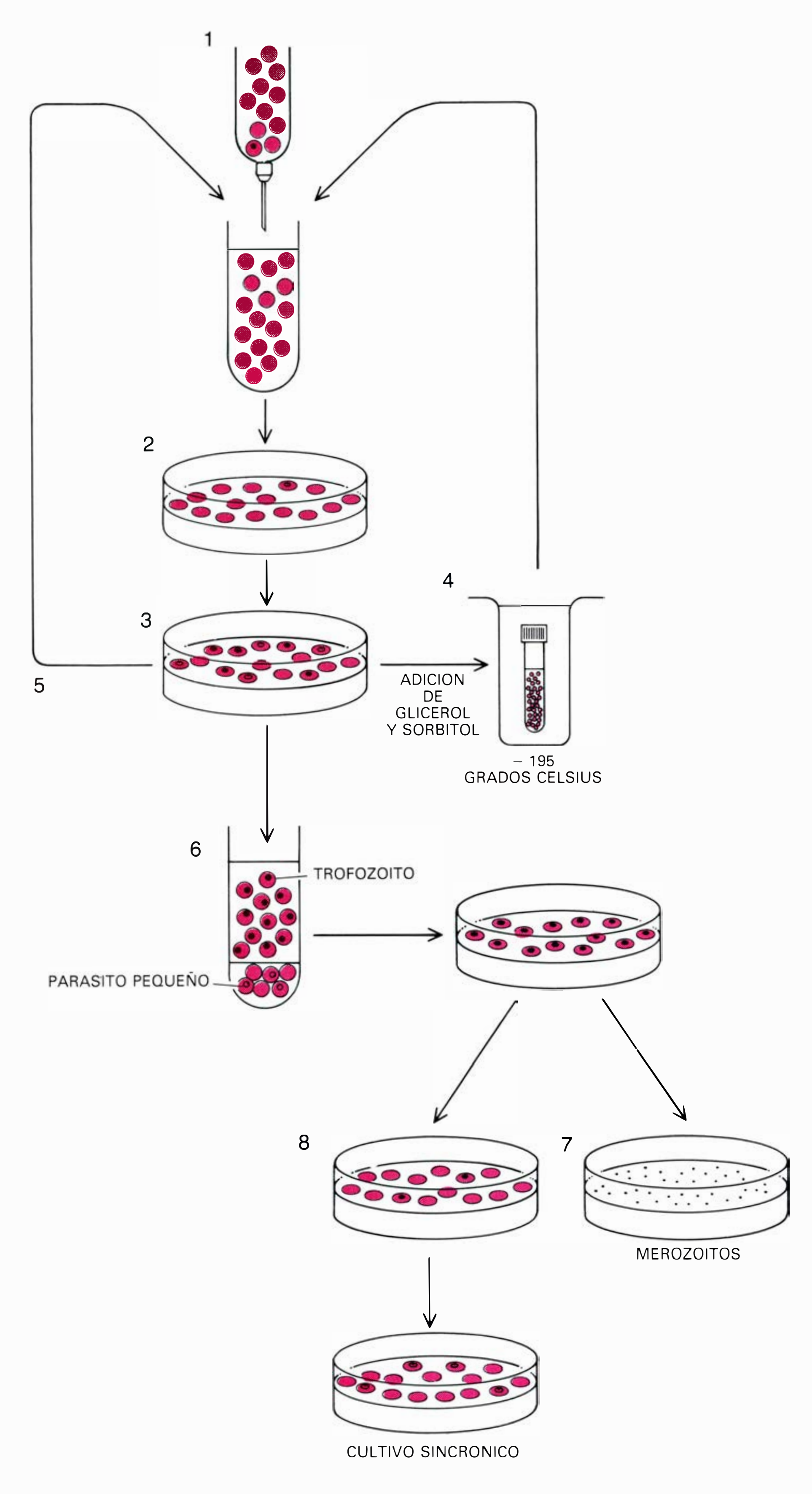
El glóbulo rojo sanguíneo, en donde el parásito de la malaria encontrará las funciones celulares alteradas, gobernadas por estos genes variantes, está en su mayor parte lleno de hemoglobina, proteína que toma el oxígeno en los pulmones y lo lleva a los tejidos. Las otras proteínas del citoplasma del eritrocito son enzimas metabólicos. Algunos de ellos catalizan la glucolisis, degradación gradual de la glucosa hasta formar lactato en un proceso en el que se sintetiza trifosfato de adenosina (ATP). Otros enzimas catalizan lo que se denomina ruta de las pentosas fosfato, que mantiene los coenzimas nicotinamida adenín dinucleótido fosfato (NADP) y glutatión en su forma reducida. El ATP es el portador universal de energía; las formas reducidas del NADP (a saber, NADPH) y del glutatión se necesitan para evitar o reparar los daños por oxidación. La membrana celular rodea a la célula y controla su conformación y su deformabilidad. Controla también el movimiento de entrada y salida de iones en la célula; en particular, mantiene, a expensas del ATP, una elevada concentración interior de potasio, en contra de la tendencia a un equilibrio resultante de la baja concentración de potasio que existe en el ambiente externo. Sobre la cara externa de la membrana, las glucoproteínas y los glucolípidos dotan a la superficie de características que, ante el ambiente, se reconocen como exclusivas.

Es en esta superficie donde el parásito de la malaria interacciona por primera vez con la célula. El parásito es un pequeño protozoo unicelular del género *Plasmodium,* del que cuatro especies causan la malaria en el hombre. La enfermedad más letal, responsable de un millón de muertes anuales entre los niños africanos, está causada por *Plasmodium falciparum.* El mosquito *Anopheles* inyecta una forma especializada del parásito en el torrente circulatorio, emigra al hígado y allí se desarrolla y divide hasta producir merozoítos, que constituyen la forma que infecta a los glóbulos rojos. Los merozoítos retor-



CULTIVO CONTINUO del parásito de la malaria *Plasmodium falciparum*, mantenido en este aparato. Los parásitos se cultivan sobre una capa fina de eritrocitos de la sangre que recubre el fondo del tubo horizontal. Sobre ella se dispone el medio nutritivo. El contenido de oxígeno del medio y otras condiciones experimentales se manipulan a través de los tubos verticales. Las células se retiran, para su examen microscópico, a través del tubo corto frontal del centro. Hasta que, en 1976, se establecieron las condiciones para cultivar el parásito, las investigaciones sobre la malaria falciparum, la forma más grave de la enfermedad, habían de realizarse con sangre de donantes humanos voluntarios o de huéspedes primates. (Fotografía de James B. Jensen.)

SEGUIMIENTO DE LA ACCION del parásito de la malaria a través de frotis, sobre portaobjetos, de células cultivadas. Por tinción (Giemsa) se colorea el núcleo en púrpura oscuro y el citoplasma en azul. La tasa de multiplicación se mide contando el número de parásitos por cada 100 eritrocitos. Este frotis muestra el parásito en diversos estadios de su ciclo biológico, del que se presenta un esquema en la ilustración de la parte inferior de la página 101.



CULTIVO DE *P. FALCIPARUM*. Se inicia inoculando células parasitadas, procedentes de un enfermo o de un cultivo de reserva, en eritrocitos humanos frescos (*1*). Las células se desarrollan a la temperatura del cuerpo (en placas de cultivo, como se muestra aquí, o en el tubo en U descrito en la página precedente) y forman una fina capa cubierta por el medio nutritivo (*2*). Pueden tomarse muestras del cultivo para medir la multiplicación del parásito. Al cabo de tres o cuatro días, las células contienen una mezcla de parásitos grandes y pequeños (*3*), que pueden congelarse para una utilización posterior (*4*), inocularse en nuevos cultivos (*5*) o incubarse en una solución de gelatina para separar los trofozoítos (*6*); los parásitos grandes pueden cultivarse aisladamente para producir merozoítos, la forma infecciosa, que se someterán luego a estudio (*7*), o mezclarse con células nuevas para producir un cultivo sincrónico del parásito (*8*).

nan al torrente sanguíneo, reconocen a la membrana de los glóbulos rojos y se fijan a ella. Se activa un mecanismo, todavía poco conocido, que permite al merozoíto introducirse en la célula cerrándose la membrana envolvente. Encerrado en una vacuola, el parásito crece digiriendo hemoglobina para adquirir los aminoácidos necesarios para fabricar sus propias proteínas y, creemos, aprovechando para su metabolismo la glucosa, el ATP y los coenzimas del eritrocito. Después de un período de crecimiento, el núcleo del parásito se divide varias veces y, seguidamente, los núcleos, con el correspondiente citoplasma, quedan rodeados de membrana. Se forman así entre 12 y 24 nuevos merozoítos, que hacen estallar la célula, se salen de ella e invaden otras células.

Las fiebres periódicas, características de este mal, se deben a la liberación sincrónica, en todo el cuerpo, de merozoítos y de toxinas producidas por el parásito. En el caso de *P. falciparum*, esta liberación tiene lugar cada 48 horas, que es el período del ciclo de desarrollo del parásito en el interior del glóbulo rojo. La fiebre y el debilitamiento que la acompaña son los principales síntomas de la malaria. Sin embargo, en la malaria falciparum hay además un efecto de mayor gravedad. El eritrocito infectado desarrolla sobre su superficie unas protuberancias, que fijan la célula a las paredes de los capilares hasta que el parásito está maduro. Si muchas células se agrupan de este modo en un órgano vital, el cerebro por ejemplo, puede sobrevenir la muerte. Tenemos la indicación más clara del efecto protector del gen de la anemia falciforme en que son muy pocos los portadores del gen que mueren por complicaciones cerebrales derivadas de la malaria falciparum.

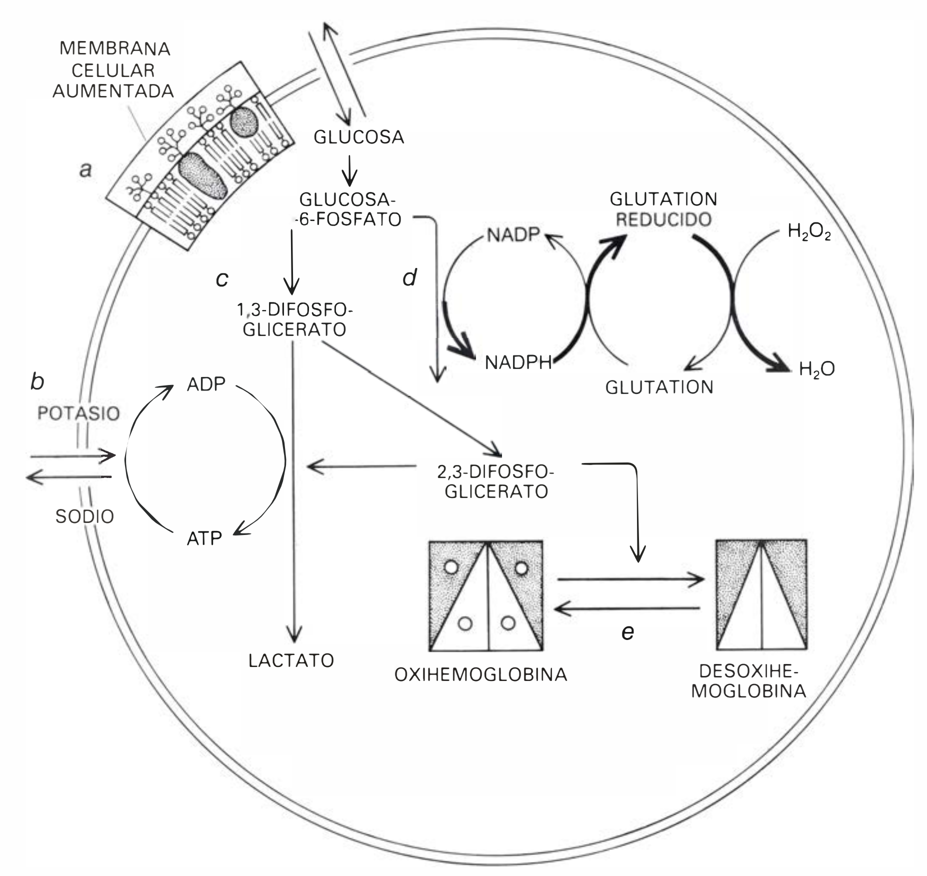
Linus Pauling y sus colegas fueron los primeros en definir una enfermedad molecular, al demostrar que en la anemia falciforme la hemoglobina estaba alterada y que las personas con esta anemia sólo presentaban en sus eritrocitos la molécula alterada, la hemoglobina *S*, mientras que algunos miembros de sus familias poseían aproximadamente una mitad de su hemoglobina del tipo *S* y la otra mitad del tipo *A*, el normal. Estudios similares, publicados por las mismas fechas, confirmaron los hallazgos de Pauling; mostraban que el patrón de herencia de la anemia falciforme podía adscribirse a un solo gen, y que la enfermedad aparecía sólo en los miembros de la familia que eran homozigóticos para este gen.

MALARIA
GEN PARA LA CELULA FALCIFORME
GEN PARA LA BETA-TALASEMIA

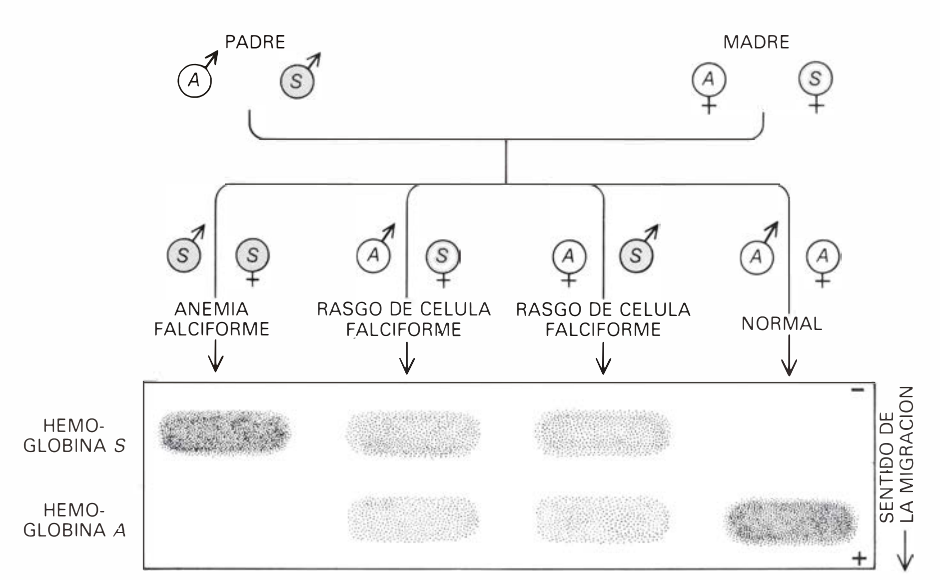
DISTRIBUCION del gen de la célula falciforme y del gen de la beta-talasemia. Coincide con aquellas zonas donde la malaria falciparum era habitual antes de 1930 (*color*). Esta concordancia geográfica proporcionó la primera indicación de que la resistencia a la malaria podía ser la ventaja evolutiva que tendía a mantener, en ciertas poblaciones humanas, una elevada frecuencia de genes responsables de la persistencia de algunas enfermedades sanguíneas.

GLOBULO ROJO DE LA SANGRE
1
2
3
TROFOZOITO
4
5
6
MEROZOITOS
7
GAMETOCITO FEMENINO
GAMETOCITO MASCULINO

CICLO BIOLOGICO DE *P. falciparum* en el eritrocito. Se inicia con la invasión de la célula por un merozoíto (*1*). El parásito engloba una gotita de citoplasma, de suerte que, en sección, parece un pequeño anillo (*2*). El anillo crece y se rellena hasta convertirse en un trofozoíto (*3*); sobre la superficie de la célula se desarrollan estructuras en forma de protuberancias que fijan la célula a la pared del vaso sanguíneo. El núcleo del parásito se divide repetidamente (*4*); cada núcleo hijo adquiere una fracción de citoplasma (*5*); el parásito se divide en 12 o 14 merozoítos, que hacen estallar la célula y comienzan un nuevo ciclo (*6*). Algunos trofozoítos se desarrollan (*7*), dando gametocitos macho y hembra que, ingenidos por mosquitos, iniciarán la fase sexual.



EL ERITROCITO, la célula más sencilla de la sangre, no es más que citoplasma rodeado por una membrana. Los glucolípidos y las glucoproteínas localizados sobre una superficie (*a*) determinan las interacciones de la célula con su entorno y con el parásito de la malaria. Las proteínas de la membrana controlan el transporte de sustancias hacia el interior y el exterior de la célula; en particular, el potasio se bombea hacia dentro y el sodio hacia fuera (*b*). Este proceso requiere energía, que proporciona el ATP generado a través de la glucolisis, proceso por el cual la glucosa se degrada hasta formar lactato (*c*). La ruta de las pentosas fosfato (*d*) produce NADP reducido y glutatión reducido; ambos evitan y reparan el daño de los agentes oxidantes en la membrana celular. El citoplasma contiene, en una gran proporción, la proteína hemoglobina, formada por cuatro cadenas (*e*). La hemoglobina fija oxígeno en el pulmón y lo cede a los tejidos, sufriendo durante este proceso cambios estructurales que, en cierta medida, están controlados por el difosfoglicerato. La carencia de núcleo priva al glóbulo rojo de muchas funciones.



HERENCIA de los genes para la hemoglobina *S* anormal, esquematizada para el caso de padres heterozigóticos para el gen anormal. Un hijo es homozigótico (*SS*) para el gen, es decir, que hereda dos genes para hemoglobina *SS* y padece anemia falciforme. Dos hijos heredan un gen para la hemoglobina *S* y otro para la hemoglobina normal, *A:* son heterozigóticos (*AS*), como sus padres, y presentan el "rasgo" de célula falciforme. Un hijo es normal (*AA*). Las hemoglobinas se identifican por electroforesis (*abajo*). Las muestras de hemoglobina, colocadas sobre acetato de celulosa, se someten a una corriente eléctrica. La hemoglobina *S* difiere de la hemoglobina *A* en un solo aminoácido; el aminoácido sustituido carece de una carga negativa y por ello la hemoglobina *S* emigra hacia el polo positivo a menor velocidad que la hemoglobina *A*. Los heterozigotos *AS* (*bandas centrales*) poseen una mezcla de las dos hemoglobinas.

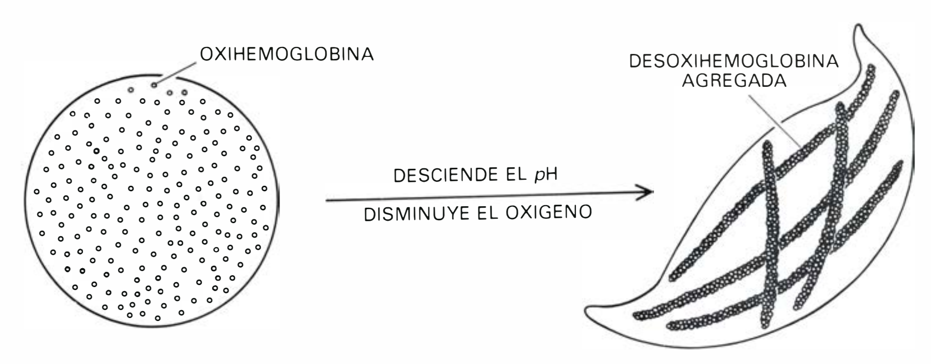
La hemoglobina del adulto consta de dos cadenas alfa y dos cadenas beta. Vernon M. Ingram, de la Universidad de Cambridge, comprobó muy pronto que, en la hemoglobina *S*, sólo la cadena beta era anormal. La anormalidad atañía exclusivamente a uno de los aminoácidos constituyentes de la cadena: una valina había sustituido a un ácido glutámico. Los homozigotos respecto a la anemia falciforme (designados por *SS*) sólo disponen de hemoglobina *S*, pues son portadores de dos genes mutantes de la cadena beta. Los heterozigotos (*AS*), de los que se dice que tienen el "rasgo" o "carácter" de célula falciforme, llevan sólo un gen mutante de la cadena beta y aproximadamente el 40 por ciento de su hemoglobina es hemoglobina *S*. Si dos personas con el rasgo de célula falciforme tienen cuatro hijos, lo probable, de acuerdo con los principios mendelianos, es que un hijo padezca de anemia falciforme, otro sea normal y los dos restantes posean el rasgo de célula falciforme.

Los síntomas de la anemia falciforme aparecen cuando los eritrocitos *SS* pierden oxígeno al circular por los tejidos. Cuando la hemoglobina *S* se desoxigena, tiende a agregarse formando fibras largas y delgadas. Estas fibras distorsionan las células, que normalmente tienen forma de disco, y les confieren forma de media luna. En una crisis de anemia falciforme, algunas células falciformes bloquean la circulación local de la sangre e impiden la llegada del oxígeno. A medida que baja el nivel de oxígeno, más eritrocitos adoptan la forma de hoz y se agranda la zona de mala circulación, causando la muerte extensiva del tejido. En ausencia de un tratamiento médico avanzado, la supervivencia de los homozigotos para la hemoglobina *S* es muy baja. Las personas con el rasgo de célula falciforme no suelen, sin embargo, sufrir la enfermedad. Sus eritrocitos *AS* poseen suficiente hemoglobina normal y sólo adoptan la conformación falciforme bajo condiciones extremas, tales como una elevada altitud.

Hace cuatro años, tras cincuenta de esfuerzos por cultivar el parásito de la malaria en el laboratorio, uno de nosotros (Trager) logró definir las condiciones que permitían mantener a *P. falciparum* en un cultivo continuo de eritrocitos humanos en un medio artificial semejante a la sangre. El sistema de cultivo lo están explotando algunos investigadores para desarrollar vacunas experimentales que contienen material de los distintos estadios del ciclo biológico del parásito; se están ensayando en

animales. Otros investigadores pretenden purificar y analizar los agentes bioquímicos que tienen efectos particulares en el curso de una infección. El sistema nos ha permitido también estudiar en detalle las interacciones del parásito de la malaria con eritrocitos anormales.

Para saber cómo protege la hemoglobina de una célula falciforme a un portador heterozigótico frente a la malaria, cultivamos el parásito en glóbulos rojos de sangre procedente de donantes normales, de individuos con el rasgo de células falciforme y de pacientes con anemia falciforme. Realizamos el ensayo bajo condiciones de cultivo normalizadas, en una atmósfera con una concentración del 17 por ciento de oxígeno, que creaba una tensión de oxígeno en el medio de cultivo similar a la de los pulmones. En esta situación, la hemoglobina aparecía totalmente oxigenada y las células variantes no se hacían falciformes. El parásito crecía igualmente bien en los tres tipos de células. Se demostraba así que, en las células variantes, no había ninguna alteración importante del metabolismo del eritrocito y que la hemoglobina *S*, como la hemoglobina *A*, podía ser digerida por el plasmodio.

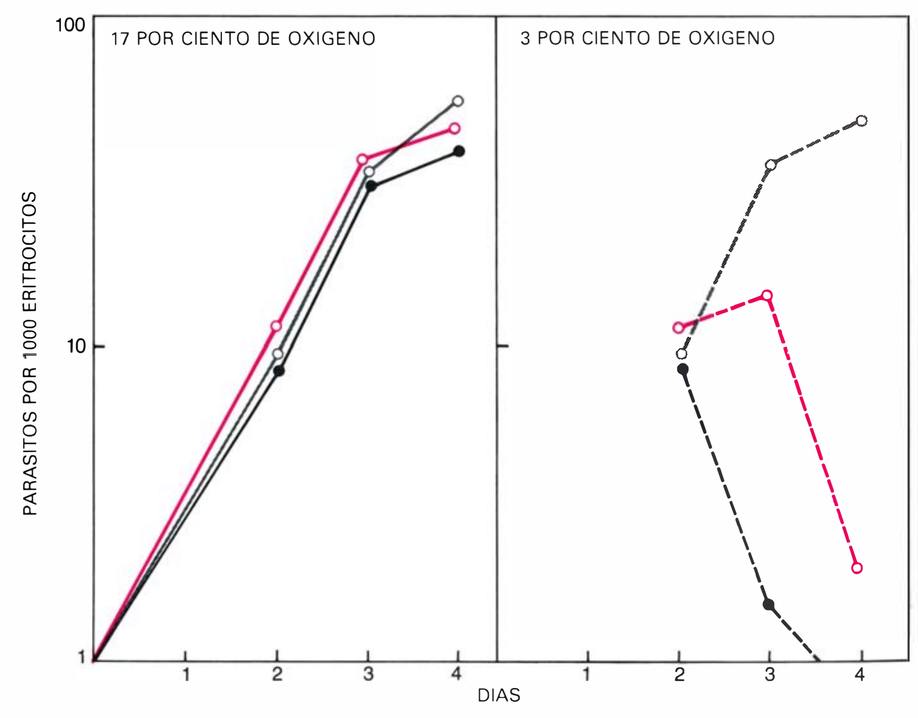


HEMOGLOBINA *S* dispersa por todo el eritrocito, en su estado desoxigenado (*a la izquierda*); posee una forma normal de disco. A diferencia de la hemoglobina normal, *A*, la *S* tiende a agregarse cuando queda desoxigenada en los tejidos y forma estructuras aciculares, casi cristalinas, que distorsionan la célula confiriéndole una forma rígida (*a la derecha*). Las células falciformes pueden bloquear capilares, lo que, a su vez, disminuye el riego, reduce el nivel de oxígeno e induce a otras células a adoptar la forma de hoz.

Para ensayar el efecto de la forma de hoz sobre el crecimiento del parásito, añadimos un pequeño número de células normales infectadas a cultivos de células *SS* y *AS*, en una atmósfera con el 17 por ciento de oxígeno. Durante las 48 horas siguientes, el período de un ciclo de crecimiento, todos los parásitos abandonaron sus células huésped e invadieron las células variantes. Cuando bajamos la concentración de oxígeno hasta el 3 por ciento, las células *SS* se hicieron falciformes, al igual que algunas de las células *AS*. Controlábamos la situación del parásito diariamente. Al cabo de una jornada con baja tensión de oxígeno, casi no se apreciaban parásitos en los cultivos de células *SS*. Se habían lisado o desintegrado, y eso mismo habían hecho sus células huésped. En los cultivos de células *AS*, por el contrario, sólo después del segundo día disminuyó de forma significativa el número de parásitos vivos y, más que desintegrarse, los parásitos muertos semejaban masas arrugadas dentro de las células. Nos daba la impresión de que los parásitos se habían agotado por inanición, tal vez como consecuencia de algún tipo de inhibición metabólica, resultante, indirectamente, de la morfología falciforme.

Si este era el caso, evitar que las células adoptasen la forma de hoz debería proteger a los parásitos. Tratamos células *AS* con cianato, que incrementa la afinidad de la hemoglobina *S* por el oxígeno, haciendo menos probable su agregación a una determinada tensión de oxígeno. Después del tratamiento con cianato y un lavado, las células *AS* seguían siendo huéspedes viables para *P. falciparum*, pero ahora ya no se tornaban falciformes con tanta facilidad. Si se infectaban estas células y se cultivaban a un 3 por ciento de oxígeno, los parásitos sobrevivían. La inhibición observada en células sin tratar debía derivarse, por tanto, de la adopción de la forma en hoz de las células y no simplemente de la baja tensión de oxígeno. ¿A través de qué mecanismo podría la configuración falciforme inhibir el crecimiento del parásito?

Cuando un eritrocito que contiene hemoglobina *S* adopta la forma de hoz, la membrana celular se hace más permeable al potasio, que se fuga al exterior; en condiciones de baja tensión de oxígeno, el nivel de potasio en nuestras células huésped *AS* había disminuido. Se había visto anteriormente que los parásitos mantenidos fuera de los glóbulos rojos requerían un ambiente con elevada concentración de potasio y, por ello, planteamos la hipótesis de que la pérdida de potasio a lo largo de la formación de células falciformes podría inhibir el metabolismo del parásito. Para probarlo, incubamos de nuevo células *AS* infectadas en un medio con un 3 por ciento de oxígeno y elevado contenido de potasio. Las células se volvieron falciformes, como de costum-



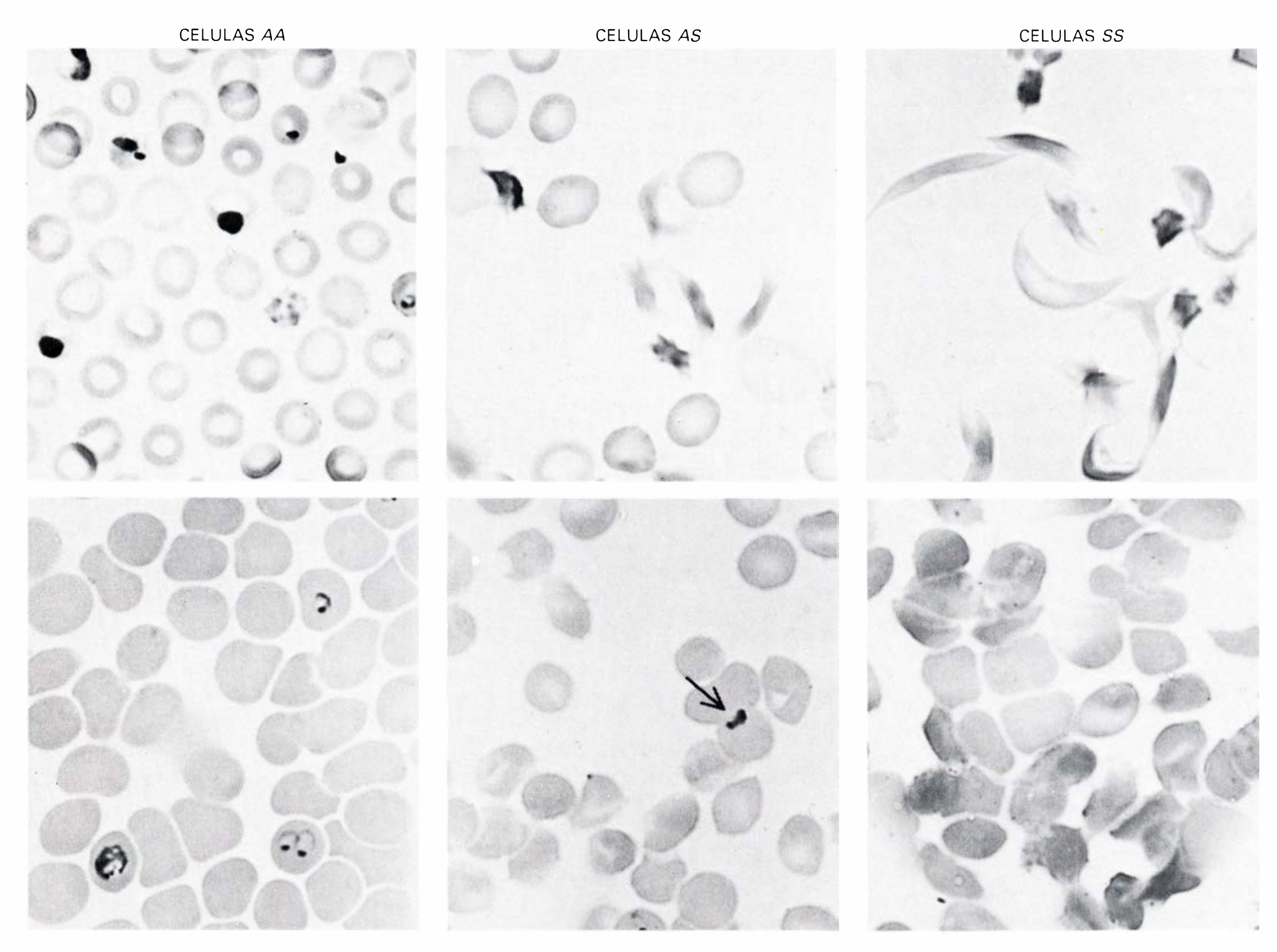
MULTIPLICACION DE *P. falciparum*, que viene a coincidir en células normales (*círculos vacíos*), en células *AS* (*color*) y en células *SS* (*negro*), siempre que las células estén en un medio con un 17 por ciento de oxígeno (*izquierda*). Pero si a los dos días se baja la concentración de oxígeno hasta el 3 por ciento, solamente los parásitos que están en células normales continuarán creciendo (*derecha*). Los parásitos localizados en células *SS* mueren al cabo de un día y los que se encuentran en células *AS*, en dos días.

bre, a una tensión de oxígeno baja, pero ahora el nivel de potasio celular permanecía alto y los parásitos sobrevivieron. Dicho de otra manera, al obviar la pérdida de potasio, se evitaba la inhibición del crecimiento del parásito en las células falciformes *AS*. (Bajo las mismas condiciones, los plasmodios localizados en células *SS* falciformes no quedaban protegidos: morían por lisis. Las micrografías electrónicas mostraron por qué. Después de seis horas de desoxigenación, podían verse unos haces aciculares de hemoglobina *S* agregada, que penetraban en algunos de los plasmodios. Las membranas de otros parásitos se habían desgarrado y estaban parcialmente lisadas. En otras palabras, en las células *SS*, los parásitos no morían por inhibición metabólica, sino por destrucción física.)

La secuencia de acontecimientos que ocurría en las células *AS* parecía, por tanto, ser la siguiente: el paso a células falciformes disminuía el nivel de potasio y el bajo nivel de potasio mataba a los parásitos. Sin embargo, este hallazgo no podía explicar, por sí solo, la resistencia de los heterozigotos a la malaria. Puesto que una célula *AS* sin infectar posee menos hemoglobina *S* que una célula *SS*, no suele hacerse falciforme en la naturaleza; circula a través de las regiones de baja tensión de oxígeno demasiado de prisa como para que se convierta en falciforme. Sin embargo, el avance de una célula parasitada se ve impedido por las protuberancias de su superficie y la célula permanece durante muchas horas en un ambiente con baja tensión de oxígeno. Incluso así, se harían falciformes menos del 5 por ciento de las células, a no ser por algún otro efecto de la infección.

Lucio Luzzatto y sus colegas, del Istituto Internazionale di Genetica e Biofisica de Nápoles, han demostrado que las células infectadas se hacen falciformes mucho más de prisa que las células sin infectar. ¿Por qué? Descubrimos que el ambiente intracelular de una célula infectada era más ácido (0,4 unidades de *p*H más bajo) que el de la célula sin infectar y que la tasa de paso a la forma de hoz aumentaba significativamente. Calculamos, además, que el bajo *p*H de las células parasitadas aumentaba también la cuantía de la deformación hasta en un 40 por ciento.

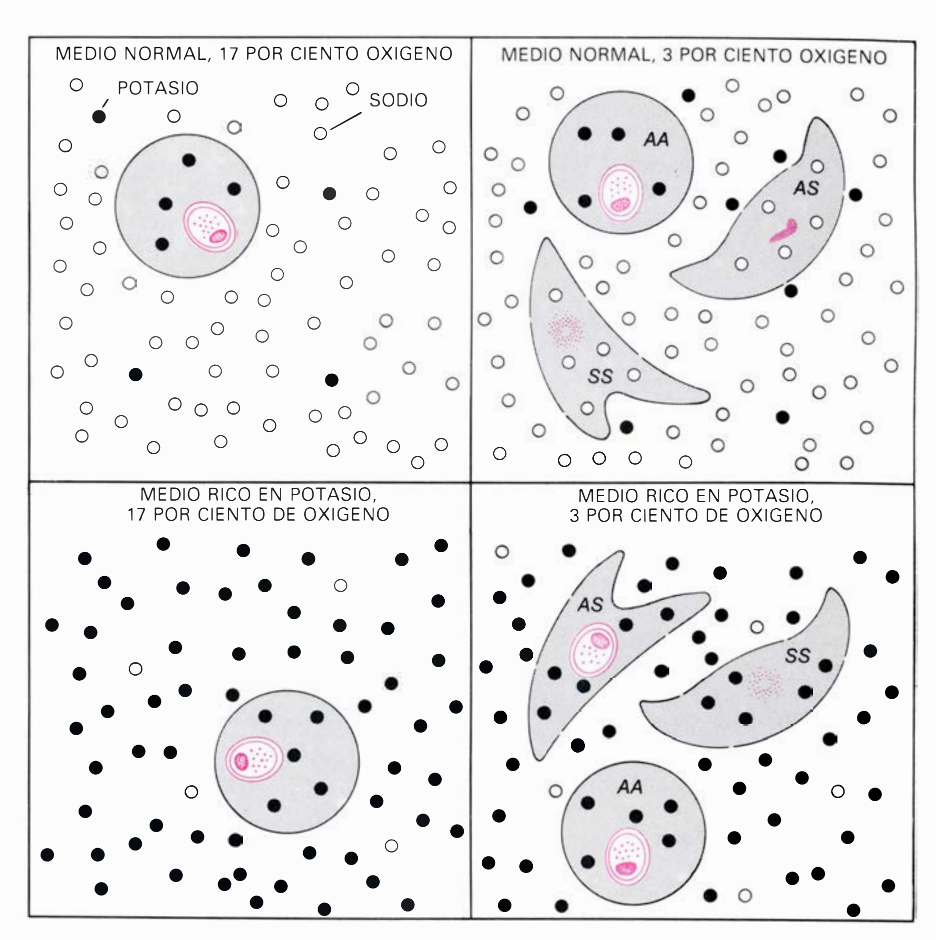
En conjunto, estas observaciones sugieren el siguiente mecanismo de protección contra la malaria en las células falciformes heterozigóticas. El parásito en una célula *AS* infectada se comporta normalmente hasta que la célula queda retenida en un tejido. Entonces, dada la baja tensión de oxígeno del ambiente y el bajo *p*H intracelular, la célula huésped se hace falciforme. El nivel de potasio desciende y el parásito muere. Tal proceso puede proteger contra la malaria, aun cuando no todos los parásitos se vean afectados, porque hasta una reducción en la tasa de multiplicación del plasmodio puede dar al sistema inmunitario el tiempo que necesita para montar su propia respuesta protectora. (Existe una hipótesis alternativa. Las células infectadas podrían, por alguna razón que se nos escapa, tornarse falci-



ERITROCITOS INFECTADOS CON EL PARASITO, cultivados durante un día en una atmósfera con un 3 por ciento de oxígeno. Las preparaciones superiores se fijaron con formaldehído para conservar la forma de las células; las inferiores son frotis sobre portaobjetos, donde se distinguen mejor los parásitos. Las células normales (*AA*) (*a la izquierda*) mantienen la forma de disco; algunas células *AS* (*en el centro*) y todas las células *SS* (*a la derecha*) han adoptado una configuración falciforme. Los parásitos de las células *AA* viven. En las *AS* los parásitos están muertos (*flecha*) o morirán antes de 24 horas.

formes mientras circulan, en lugar de hacerlo al ser secuestradas, y podrían entonces ser eliminadas por la acción filtrante del bazo. La primera hipótesis, sin embargo, viene apoyada por la evidencia de que los heterozigotos no están protegidos contra los tipos de malaria en los que las células infectadas no desarrollan protuberancias ni quedan secuestradas en los tejidos.)

Además de la anemia falciforme y el rasgo de célula falciforme, existen otros trastornos hereditarios de los eritrocitos cuya incidencia geográfica se ha correlacionado con la de la malaria, lo que implica que los genes responsables de estos trastornos pueden conferir también alguna resistencia. Entre ellos se cuentan las talasemias, que comportan una deficiencia en la fabricación de una u otra de las cadenas de la hemoglobina. La beta-talasemia, por ejemplo, es una deficiencia en la síntesis de la cadena beta. La talasemia beta homozigótica, conocida como anemia de Cooley, constituye una grave enfermedad en la que la síntesis de hemoglobina de adulto normal es escasa, si no nula. La transfusión sanguínea suele ser el único medio de prolongar la vida del paciente. Sin embargo, en muchas regiones afectadas por la malaria, en particular en torno a las costas del Mediterráneo, aproximadamente el 1 por ciento de los niños que nacen son homozigóticos para el gen de la beta-talasemia y padecen anemia de Cooley; los heterozigóticos se libran de la enfermedad. Aunque la resistencia a la malaria no se ha demostrado de forma convincente en los heterozigotos beta-talasémicos, existe una correlación geográfica muy sugestiva entre la frecuencia del gen y una historia regional de malaria. Por citar sólo un ejemplo, la frecuencia del gen es mucho más elevada en los valles de Cerdeña, donde la malaria fue endémica durante mucho tiempo, que en las montañas, donde era más rara.



SE HA DEMOSTRADO QUE LA PERDIDA DE POTASIO causa la muerte del parásito en células *AS*. En una atmósfera con un 17 por ciento de oxígeno (*izquierda*), la membrana celular permanece intacta y el nivel de potasio se mantiene adecuadamente, tanto en un medio fisiológico normal (con poco potasio) (*arriba*) como con elevada concentración de potasio (*abajo*). En un medio con baja tensión de oxígeno y baja concentración de potasio (*arriba, derecha*) las células *AS* y *SS* adoptan la forma en hoz; sus membranas se rompen, pierden potasio y sus parásitos se colapsan (*AS*) o desaparecen (*SS*). La incubación en un medio con elevada concentración de potasio (*abajo, derecha*) protege a los parásitos en las células *AS*, aun cuando adquieran una configuración falciforme y se rompa la membrana, pero no protege a los parásitos que están en células *SS*. La causa de la muerte del parásito en células *AS* es, por tanto, la privación de un factor metabólico esencial, el potasio. En células *SS* la muerte sobreviene por penetración física.

Nos propusimos demostrar un efecto de resistencia en nuestro sistema de cultivo. Sabíamos que una de las características de las células talasémicas era una sensibilidad anormal de su membrana celular a la oxidación. Cuando una molécula se oxida, se eliminan electrones que ordinariamente desempeñarían un papel en la formación de enlaces químicos. Cuando las moléculas lipídicas (grasas) se oxidan, se degradan y destruyen la integridad de la membrana. No se han identificado todos los agentes de oxidación que hay en las células, pero se sabe que el peróxido de hidrógeno es uno de ellos.

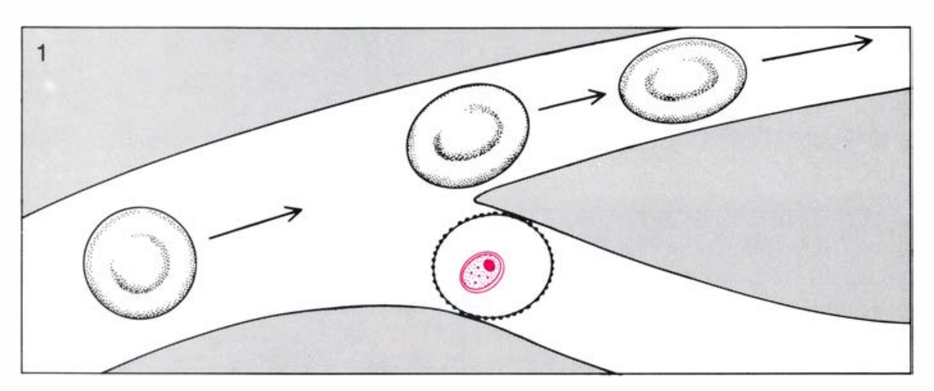
El parásito de la malaria genera peróxido de hidrógeno en su célula huésped (como demostraron N. Etkin y John W. Eaton, de la Facultad de Medicina de la Universidad de Minnesota). Los peróxidos originan en las células una tensión oxidante que pone en peligro la capacidad de la célula para conservar su integridad. En los eritrocitos talasémicos más sensibles, el peróxido de hidrógeno podría dañar la membrana. Confirmamos la verosimilitud de este efecto al descubrir que los parásitos localizados en las células heterozigóticas para la talasemia beta eran más sensibles a tres condiciones experimentales que los parásitos de células normales. Una de las situaciones de ensayo era un ambiente con elevada presión de oxígeno (una concentración de oxígeno entre el 25 y el 30 por ciento). Otra condición era la presencia de ciertos productos químicos que catalizan reacciones de oxidación. La tercera se refería a la ausencia, en el medio de cultivo, de uno de los constituyentes normales, la forma reducida del glutatión, que es intermediario en la ruta metabólica que rebaja el nivel de peróxido de hidrógeno en la célula.

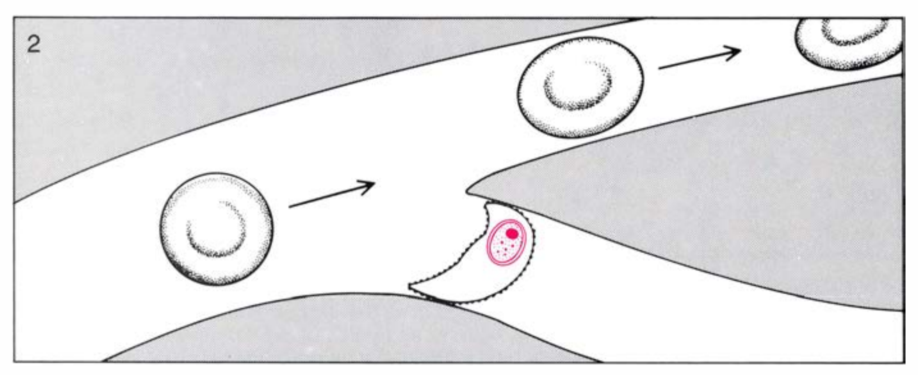
Se habían ideado las tres condiciones para incrementar la tensión oxidante en la célula. Eaton comprobó que, en ratones infectados con malaria, la tensión oxidante y la protección están finamente equilibradas. Cualquier sensibilidad adicional en una célula talasémica podría, por tanto, afectar el curso de la infección. (Aunque la célula talasémica manifiesta un bajo contenido de hemoglobina, parece bastar esa cantidad para mantener el parásito, pues el bajo nivel de hemoglobina, por sí solo, no inhibía la multiplicación del parásito en nuestro sistema.)

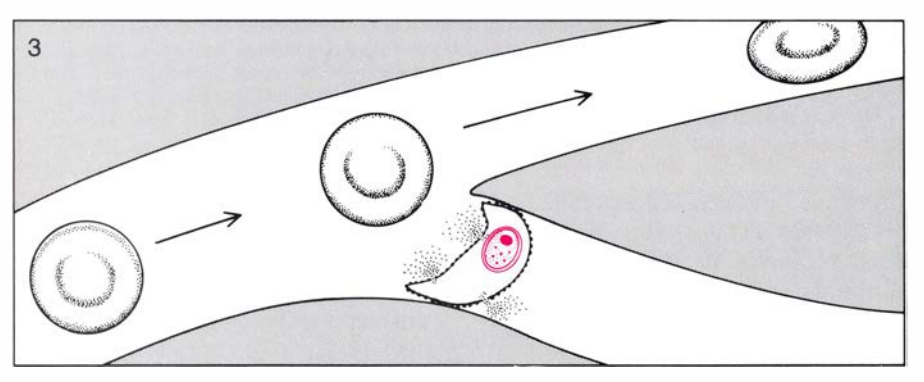
Hallamos que la vitamina E, que protege los lípidos de la membrana frente al daño por oxidación, evitaba la muerte de los parásitos en células con el rasgo de talasemia, bajo todas las condiciones. Ello constituía una prueba en pro de la idea de que la membrana de la célula heterozigótica para la tala-

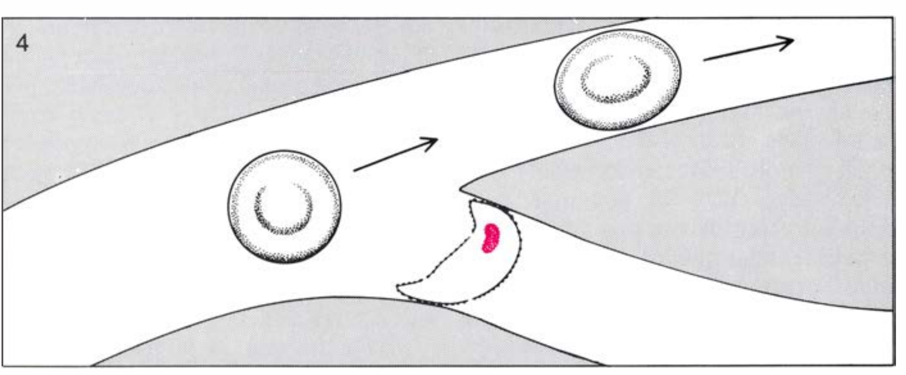
semia beta sufría daños por oxidación durante el curso de la infección malárica. Como ocurría con el mecanismo, aquí de nuevo parece estar implicado el potasio. En un medio con elevada concentración de potasio, los parásitos en células con el rasgo de talasemia se desarrollaban normalmente bajo condiciones que les hubieran conducido a la muerte en un medio con una concentración normal de potasio.

Esta investigación nos llevó a tres interesantes deducciones. Una fue la confirmación de que existía una relación, sospechada desde hacía tiempo, entre la malaria y una deficiencia hereditaria de los eritrocitos en el aporte del enzima glucosa-6-fosfato deshidrogenasa (G6PD). Esta deficiencia prevalece también en las regiones donde hay malaria. Estudios de campo han demostrado que existe a veces, pero no siempre, una correlación con la resistencia a la malaria. La G6PD es el primer enzima de la vía de las pentosas fosfato que regenera NADPH, un coenzima esencial para la protección frente al daño por oxidación y para su reparación. Resultaba que los eritrocitos deficientes en G6PD, como los del rasgo de la talasemia, podrían ser más sensibles al peróxido de hidrógeno generado por el parásito de la malaria. Nosotros encontramos que los parásitos localizados en células deficientes en G6PD eran ciertamente muy sensibles a la tensión de los agentes oxidantes y estaban protegidos por agentes antioxidantes.









CADENA ESQUEMATICA DE ACONTECIMIENTOS que protegen a los heterozigotos *AS* contra la malaria. Una célula infectada por el parásito se caracteriza por presentar protuberancias sobre su superficie y por su bajo *p*H intracelular. Una célula *AS* sin infectar pasará por un corto período de baja oxigenación dentro de un capilar, sin hacerse falciforme (*1*); por el contrario, una célula infectada quedará secuestrada el tiempo suficiente para que el bajo nivel de oxígeno y el bajo *p*H le confieran la forma de hoz (*2*). Esta configuración hace que la membrana celular pierda potasio (*3*). Privado de potasio, el parásito muere (*4*). La muerte de alguna fracción de los parásitos infectantes puede dar al organismo heterozigótico el tiempo necesario para desarrollar una respuesta inmunológica adecuada.

Un segundo corolario que se colegía de nuestros resultados con células talasémicas guardaba relación con el favismo, una anemia hemolítica promovida por la ingestión de habas, que se consumen en toda la región mediterránea. Las habas contienen una variedad de sustancias que podrían incrementar la sensibilidad del eritrocito a los agentes oxidantes; algunas de estas sustancias tienen que ver con los catalizadores de la oxidación introducidos en algunos de nuestros experimentos. ¿Mimetizan estos experimentos el consumo de habas por personas portadoras de eritrocitos talasémicos o deficientes en G6PD? Si es así, los resultados indicarían que, comiendo habas (y tal vez otros alimentos todavía no identificados), se eleva el nivel de protección frente a la malaria en las personas heterozigóticas para estos dos trastornos eritrocitarios. Este efecto de la dieta podría explicar también la inconsistencia de los estudios sobre resistencia a la malaria efectuados entre tales heterozigóticos.

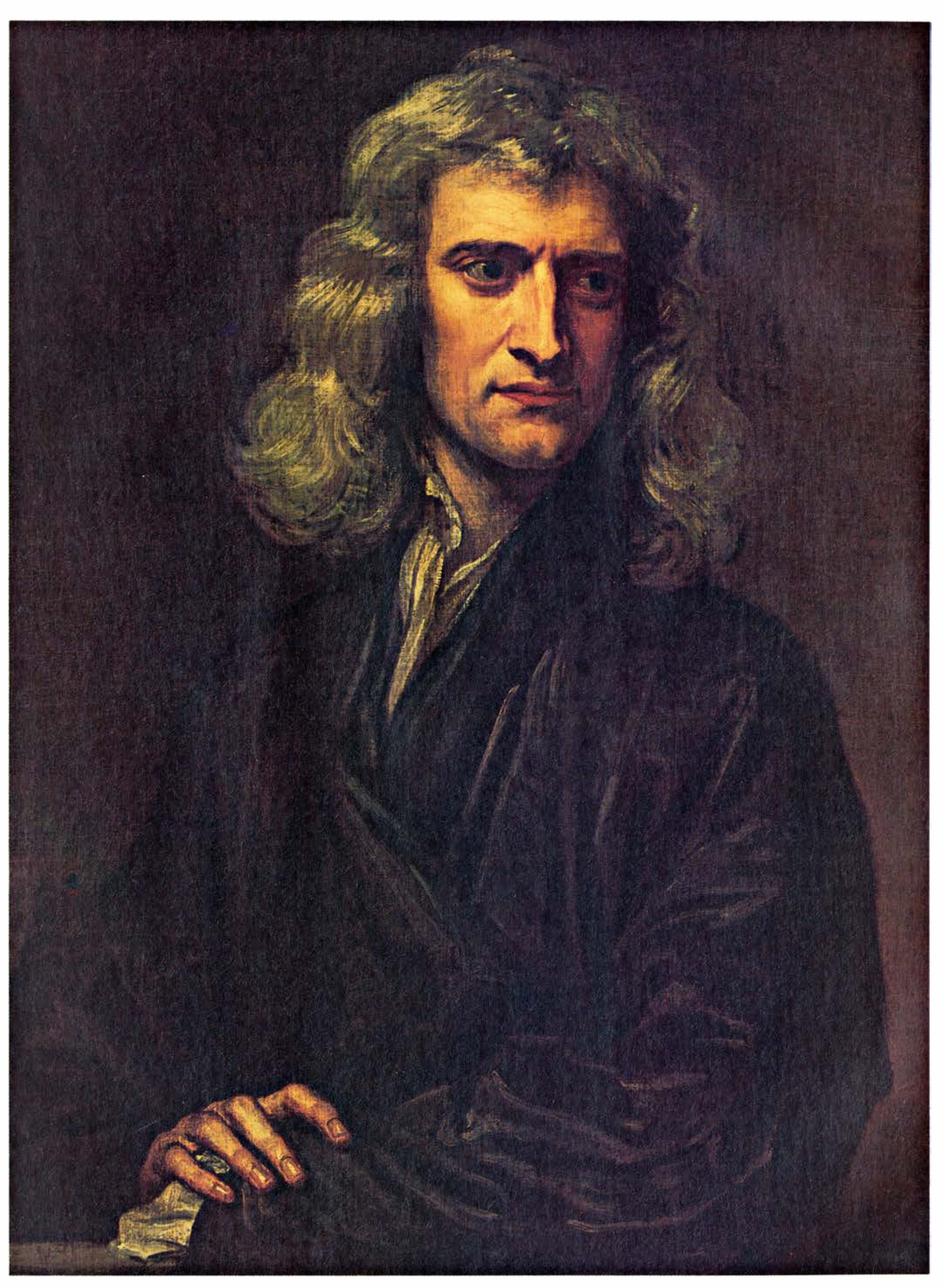
La tercera consecuencia de nuestro estudio se refiere a los niños. Durante los primeros meses de vida, los niños están protegidos, casi totalmente, contra la malaria. Geoffrey Pasvol, R. J. M. Wilson y D. J. Weatherall, de la Universidad de Oxford, demostraron hace poco que la hemoglobina fetal (que consta de cadenas alfa y gamma y persiste durante un cierto tiempo después del nacimiento y puede encontrarse en algunos eritrocitos de adultos) puede contribuir a esta protección: los mismos eritrocitos de adultos inhibían el crecimiento de los parásitos, si contenían hemoglobina fetal. Por nuestra parte, hemos encontrado que los parásitos de la malaria localizados en eritrocitos fetales, como los que están en cé-

lulas talasémicas, son altamente sensibles a la tensión oxidante. No parece, sin embargo, que la malaria haya sido una fuerza selectiva en la evolución de la hemoglobina fetal. Probablemente, la proteína fetal evolucionó bajo una presión selectiva diferente. Posee una afinidad más elevada por el oxígeno que la hemoglobina de adulto, por lo que mejora la cesión de oxígeno al feto. La sensibilidad a la oxidación del eritrocito fetal y la correspondiente resistencia a la malaria son, probablemente, efectos secundarios de una adaptación propia de la fase de desarrollo.

Para resumir, el proceso de la evolución ha dado como resultado la selección de eritrocitos genéticamente distintos que funcionan bastante bien bajo condiciones normales, aunque pueden lesionarse cuando son infectados por *P. falciparum,* daños que matan al parásito. En otras palabras, estas células son tan marginalmente viables que la infección las hace inviables e incapaces de soportar al parásito intracelular. La alteración genética, en el caso de células *AS,* es una tendencia límite a la configuración falciforme, resaltada por la presencia del parásito. La alteración en las células con el rasgo de talasemia se cifra en una mayor sensibilidad de los lípidos de la membrana al peróxido de hidrógeno generado por el parásito. En ambos casos, el efecto final es una pérdida de potasio, que inhibe el metabolismo del parásito.

El curso evolutivo de un gen puede ser muy complejo. Cuando surge una nueva mutación al azar no se puede predecir ni su efecto ni su potencial. ¿Quién hubiera acertado a predecir que ciertos genes causarían una enfermedad letal de la sangre en un homozigoto y protegerían al heterozigoto contra la muerte por malaria? La contienda delicadamente equilibrada entre el efecto selectivo de la malaria, por un lado, y la anemia falciforme y la betatalasemia, por otro, ha dado como resultado un equilibrado polimorfismo: una situación en la que la ventaja del heterozigoto acoplada a la desventaja del homozigoto mantienen un gen mutante a un nivel bajo, pero consistente.

Los mecanismos de la vida y de la muerte encuentran sus raíces en la química y en la biología molecular. Pueden explicarse, hasta cierto punto, en el laboratorio. Sin embargo, sus efectos últimos sobre la especie humana no se deciden en el laboratorio, ni siquiera en las células de los distintos seres, sino lenta e imprevisiblemente en la lucha evolutiva a través de continentes y a lo largo de milenios.



RETRATO DE ISAAC NEWTON, pintado por Godfrey Kneller en 1689, cuando el célebre físico tenía 46 años de edad. La idea de gravitación universal había sido elaborada cuatro años antes. La obra principal de Isaac Newton, *Philosophiae Naturalis Principia Mathematica*, fue publicada en 1678.

El descubrimiento newtoniano de la gravitación

¿Cómo llegó a desarrollar Newton el concepto que habría de señalar el comienzo de la ciencia moderna? Esencialmente lo hizo comparando una y otra vez el mundo real con una esquematización matemática del mismo

I. Bernard Cohen

El momento culminante de la Revolución Científica fue el descubrimiento realizado por Isaac Newton de la ley de gravitación universal: todos los objetos se atraen unos a otros con una fuerza directamente proporcional al producto de sus masas e inversamente proporcional al cuadrado de su distancia. Al someter a una sola ley matemática los fenómenos físicos más importantes del universo observable, Newton demostró que la física terrestre y la física celeste son una misma cosa. El concepto de gravitación lograba de un golpe: revelar el significado físico de las tres leyes de Johannes Kepler sobre el movimiento planetario, resolver el intrincado problema del origen de las mareas y dar cuenta de la curiosa e inexplicable observación de Galileo Galilei de que el descenso de un objeto en caída libre es independiente de su peso. Newton había alcanzado así el objetivo kepleriano de desarrollar una física basada en las causas.

El trascendental descubrimiento de la gravitación universal, que vendrá a convertirse en el paradigma de ciencia triunfante, no fue el resultado de un aislado chispazo de genio, sino que fue la culminación de una serie de ejercicios encaminados a resolver problemas. No fue fruto de la inducción, sino de deducciones lógicas y de transformación de ideas preexistentes. El descubrimiento de la gravitación pone de manifiesto lo que, a mi juicio, es característica fundamental de todas las discontinuidades en el curso de la ciencia, desde las simples innovaciones a las revoluciones más dramáticas: crear algo nuevo por transformación de ideas ya existentes.

Newton desarrolló el concepto de gravitación universal en los primeros meses de 1685, cuando tenía 42 años. Ordinariamente, los físicos hacen sus contribuciones importantes a una edad mucho más joven, pero Newton a esa edad estaba todavía en lo que él mismo llamó "los primeros años de mi vida en orden a la invención". Los documentos que me han permitido fijar la fecha del descubrimiento han hecho posible además reconstruir el proceso que llevó hasta él.

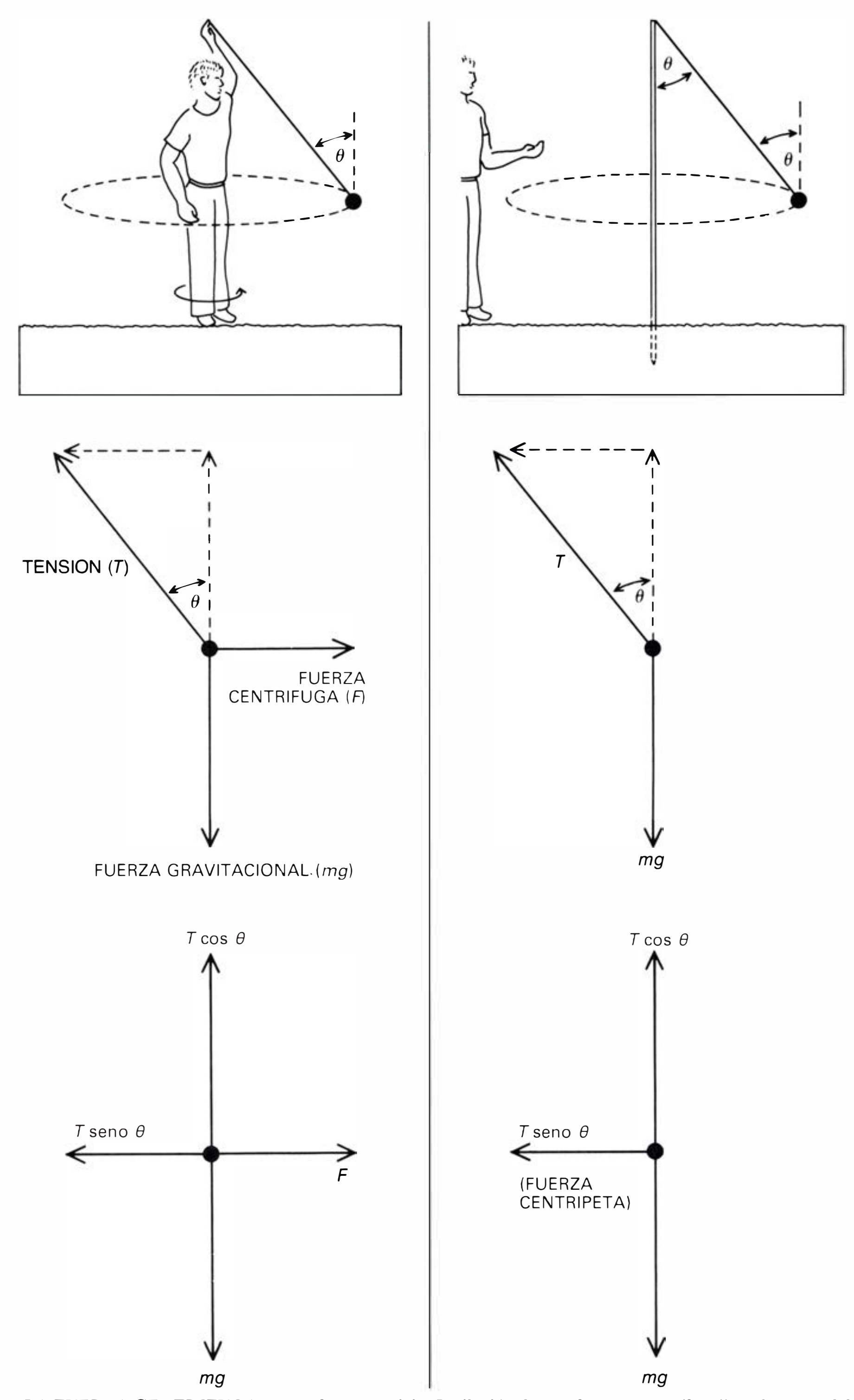
Un paso decisivo en el camino hacia la gravitación universal tuvo lugar a finales de 1679 y principios de 1680, cuando Robert Hooke introdujo a Newton en un nuevo método de analizar el movimiento a lo largo de una trayectoria curva. Hooke había tenido la habilidad de ver que el movimiento de un cuerpo que gira tiene dos componentes: una inercial y otra "centrípeta", es decir, "dirigida hacia el centro". La componente inercial tiende a lanzar el cuerpo a lo largo de una recta tangente a la curva descrita, mientras que la componente centrípeta aparta continuamente el cuerpo de esa trayectoria rectilínea inercial. En una órbita estable, como la de la luna, ambas componentes están combinadas de forma que la luna ni se aleja siguiendo un camino tangencial, ni se acerca a la tierra describiendo una espiral.

El concepto de fuerza centrípeta sustituyó la idea más antigua y confusionista de fuerza "centrífuga", es decir, "huyente del centro". René Descartes y Christiaan Huygens habían analizado el movimiento curvilíneo en términos de esa fuerza centrífuga. Por ejemplo, Descartes había estudiado el movimiento de una bola en la superficie interna de un hueco cilíndrico y, el movimiento del agua, en un caldero movido circularmente. La bola y el agua parecen huir del centro del sistema, y por eso Descartes atribuía su movimiento al influjo de una fuerza centrífuga. Para nosotros es claro que tal fuerza no existe. La interacción de los objetos físicos no da indicios de una fuerza que se aparta del centro. La ilusión de una fuerza centrífuga surge siempre que se observa un objeto en movimiento desde un sistema de referencia en rotación.

Con este cambio de perspectiva de fuerza centrífuga a fuerza centrípeta, se llegó a apreciar el papel fundamental que desempeña el cuerpo central. El análisis centrífugo había fijado su atención en el objeto que da vueltas, cuya tendencia a apartarse del centro parece no tener nada que ver con las propiedades del cuerpo central. Por el contrario, el concepto de fuerza centrípeta atiende fundamentalmente a ese cuerpo central, hacia el cual es impulsado o atraído el cuerpo que gira. Es obvio esperar que esa interacción entre el cuerpo central atrayente y el objeto circundante atraído haya representado un papel importante en cualquier teoría de la gravitación.

El análisis de Hooke sobre el movimiento curvilíneo puede parecer una consecuencia tan obvia e inmediata del principio cartesiano de inercia, que Newton no debería haber necesitado a un Hooke para instruirse en ese tema, máxime en fecha tan tardía como 1679. Newton había aceptado prácticamente el principio de inercia unos veinte años antes. Pero Newton, como Descartes y como Huygens, estaba tan enzarzado con el concepto de tendencia centrífuga, que todas esas implicaciones de la física inercial no le resultaban obvias, ni mucho menos.

Hooke escribió a Newton el 24 de noviembre de 1679, sugiriéndole entablar una correspondencia privada, de tipo "filosófico", sobre temas científicos de interés para ambos. Seis años antes habían debatido públicamente so-



LA FUERZA CENTRIFUGA es una fuerza ficticia. La ilusión de una fuerza "centrífuga", o "huyente del centro", puede producirse siempre que un cuerpo en movimiento es observado desde un sistema de referencia en rotación, como cuando hace dar vueltas a una bola, sujeta al extremo de una cuerda, un observador que gira con la misma velocidad angular que la bola (*esquemas de la izquierda*). Sobre la bola actúan dos fuerzas bien conocidas: la fuerza gravitacional (*mg*) y la tensión de la cuerda (*T*), que forma un ángulo θ con la vertical. La bola no es acelerada en la dirección vertical; por consiguiente, todas las fuerzas verticales que actúan en ella han de equilibrarse; en particular, la componente vertical de la tensión (*T* cos θ) compensa la fuerza gravitacional (*mg*). Puesto que el observador y la bola giran a la vez, la bola aparenta estar en reposo, y parece que las fuerzas horizontales deban estar también en equilibrio. En consecuencia, el observador postula una fuerza centrífuga (*F*) que compense la componente horizontal de la tensión (*T* sen θ). Sin embargo, la interacción de los objetos físicos no da ningún indicio de la existencia de tal fuerza. Resulta muy distinto el análisis de las fuerzas cuando la bola da vueltas de la misma manera, pero el observador está en reposo (*esquemas de la derecha*). En este sistema de referencia fijo, el observador ve actuar sobre la bola las mismas fuerzas verticales que veía antes. En la dirección horizontal, empero, la bola no está en reposo respecto al observador, sino que se mueve según una circunferencia. En otras palabras, la bola se acelera continuamente hacia el centro, de forma que ya no ha de esperarse un equilibrio de las fuerzas horizontales. La bola está sometida a una fuerza "centrípeta", o "dirigida al centro", fuerza que coincide con la componente horizontal de la tensión de la cuerda (*T* sen θ). De esa fuerza centrípeta da indicios la interacción entre dos objetos físicos, que son la cuerda y la bola.

bre los experimentos y teorías de Newton acerca de la dispersión de la luz en un prisma y la naturaleza del color. Hooke era sólo uno de los muchos investigadores que habían rechazado las teorías ópticas de Newton. Este había quedado tan molesto de tener que defender su trabajo, que prometió solemnemente abandonar la "filosofía" (es decir, la ciencia física) porque ésta era "una señora tan litigiosa" que quien tenga algo que ver con ella pasará el resto de su vida defendiendo sus opiniones.

Hooke, entre tanto, se había convertido en secretario de la Real Sociedad de Londres A pesar de las antiguas controversias, su carta a Newton era amistosa y afable. En ella Hooke invitaba a Newton a hacer comentarios sobre alguna de sus propias hipótesis u opiniones, especialmente sobre la noción de "componer los movimientos celestes de los planetas a partir de un movimiento rectilíneo a lo largo de la tangente y un movimiento atractivo, hacia el cuerpo central" ["compounding the celestiall motions of the planetts of a direct motion by the tangent & an attractive motion towards the central body"]. Esta frase fue, según parece, la que introdujo a Newton en la idea de descomponer el movimiento curvilíneo en una componente inercial y otra centrípeta. No hay ninguna prueba de que él ya hubiera alcanzado para el movimiento circular el nivel de comprensión que poseía Hooke. En efecto, Newton todavía hablaba con frecuencia del movimiento orbital en términos de fuerza centrífuga.

En su carta, Hooke aventuraba la sugerencia de que la fuerza centrípeta hacia el sol varía en razón inversa del cuadrado de la distancia. Hooke se quedó atascado aquí. No pudo ver las consecuencias dinámicas de la profunda visión que tenía ante sus ojos, y por consiguiente no pudo dar el salto desde el presentimiento intuitivo y el proceso conjetural a la ciencia exacta. No pudo ir más allá porque le faltaron dos cosas, el genio matemático de Newton y el prestar atención a la ley de las áreas de Kepler, que influirá de una manera decisiva en el ulterior tratamiento newtoniano de la dinámica celeste. (La ley de las áreas establece que el radio vector trazado desde el sol hasta un planeta barre áreas iguales en tiempos iguales.)

En carta del 28 de noviembre, Newton escribió a Hooke que, hasta haber leído su carta del 24, no había "ni siquiera oído hablar (que recuerde) de su hipótesis de componer los movimientos

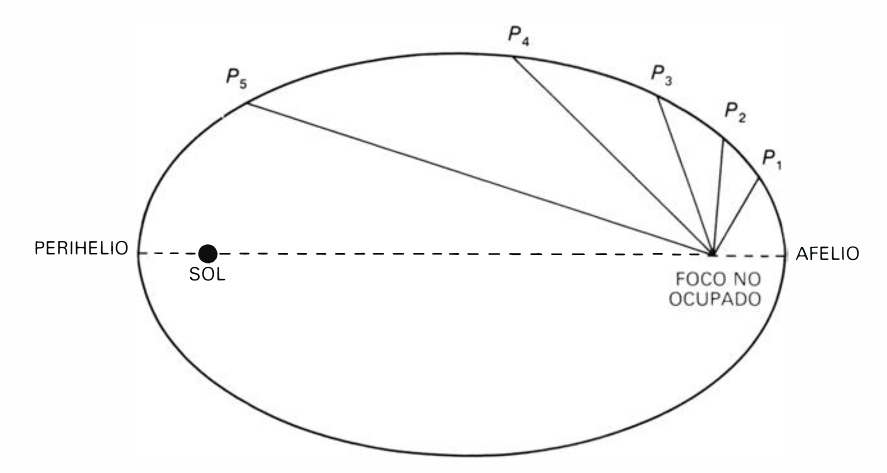
CARTA DIRIGIDA A ISAAC NEWTON por Robert Hooke, en la que incluye sus ideas sobre el análisis del movimiento a lo largo de una trayectoria curva. (Al pie de la carta, a la izquierda, puede leerse la fecha "6 de enero de 1679", escrita de acuerdo con una versión del sistema de calendario juliano para la que el año comenzaba en marzo. Según el sistema moderno de calendario, esa fecha debe leerse "6 de enero de 1680·) En la segunda frase de la carta, Hooke propone que "la atracción es siempre inversamente proporcional al cuadrado de la distancia al centro" ("the Attraction is always in a duplicate proportion to the Distance from the Center Reciprocall"). Como consecuencia, "la velocidad será proporcional a la raíz cuadrada de la atracción, y por tanto estará en razón inversa de la distancia, según supone Kepler" ("the Velocity will be in a subduplicate proportion to the Attraction and consequently as Kepler supposes Reciprocall to the Distance"). Hooke afirma que este análisis explica "todas las apariencias de los cielos". Y subraya asimismo la importancia de "encontrar las propiedades" de las curvas, porque longitudes que son "de gran interés para la humanidad" pueden deducirse del movimiento curvilíneo de la luna.

celestes de los planetas a partir de un movimiento rectilíneo a lo largo de la tangente a la curva" y un movimiento "atractivo" hacia el sol ["so much as heare (that I remember) of your Hypotheses of compounding the celestial motions of the Planets of a direct motion by the tangent to the curve..."]. Después de admitir que el análisis de Hooke era nuevo para él, Newton cambió inmediatamente de tema, proponiendo una ocurrencia de su propia fantasía: el efecto de la rotación de la tierra en la caída libre de un objeto. ¿Qué camino seguiría un objeto, dejado caer libremente, si pudiese atravesar la tierra en rotación? Newton había concluido falsamente que seguiría una trayectoria espiral.

En su carta siguiente, fechada a 9 de diciembre, Hooke detectó el error de Newton y le señaló que el camino "se parecería a una elipse" ["would resemble an Elleipse"]. Hooke deseaba que Newton se metiera en el problema del movimiento planetario, y por eso le sugirió que la descripción correcta de la caída de un objeto a través de la tierra y su propio análisis del movimiento planetario eran dos casos de "movimientos circulares compuestos de un movimiento rectilíneo y otro atractivo hacia un centro" ["circular motions compounded by a direct motion and an attractive one to a center"].

El 13 de diciembre de 1679 Newton respondió con cautela a la corrección de Hooke, pero no hizo comentario alguno sobre el análisis del movimiento circular que le proponía. Hooke no cejó en su empeño. En una carta del 6 de enero de 1680, volvió a su tesis sobre el movimiento curvilíneo y repitió la suposición cuantitativa de que la atracción centrípeta era inversamente proporcional al cuadrado de la distancia. A partir de esta suposición, concluía Hooke que la velocidad del cuerpo que gira era inversamente proporcional a su distancia al centro. Y luego señalaba que su análisis "lograba descifrar todas las apariencias de los cielos de modo muy inteligible y verdadero". Newton no contestó a esta carta.

El 17 de enero Hooke envió una breve carta suplementaria en la que escribía: "Nos queda ahora por conocer las propiedades de una *línea* curva (no circular ni concéntrica) generada por un poder atractivo central que produce las velocidades de *caída* desde la línea tangente, o movimiento rectilíneo y uniforme, tomándole a todas las *distancias* en proporción *cuadrática* inversa" ["It now remaines to know the proprietys of a curve Line (not circular nor concentricall) made by a centrall attractive power which makes the velocitys of Descent from the tangent Line or equall straight motion at all Distances in a Duplicate proportion reciprocally taken"]. En terminología moderna, podemos parafrasear así el problema de Hooke: cuando una fuerza atractiva central hace que un objeto caiga, desviándose de su trayectoria inercial y siguiendo una trayectoria curvilínea, ¿qué tipo de curva resulta, si la fuerza atractiva varía en proporción inversa al cuadrado de la distancia? Hooke concluye: "No dudo que usted, con su excelente método, encontrará fácilmente cuál ha de ser esta *curva* y cuáles sus propiedades, y sugerirá una *razón* física de esta proporción" ["I doubt not but that by your excellent method you will easily find out what that Curve must be, and its



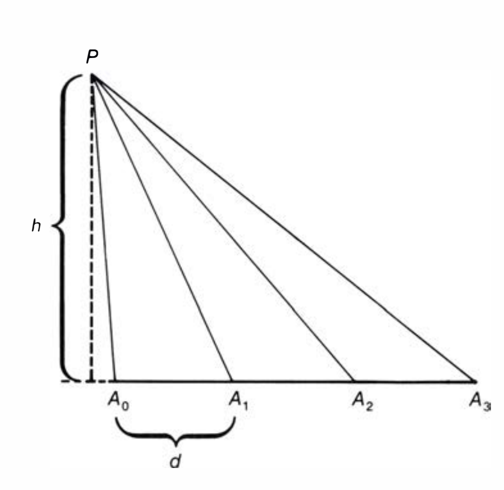
LAS POSICIONES PLANETARIAS no solían hallarse, en el siglo XVII, mediante la ley kepleriana de las áreas, sino mediante una construcción basada en considerar un radio vector que partiera del foco no ocupado de la órbita elíptica del planeta y que girase con velocidad angular uniforme. La posición (P_1, P_2, P_3, P_4, P_5) que ocupa el planeta en instantes sucesivos es la intersección de la elipse y el radio vector. La ley kepleriana de las áreas establece que el radio vector trazado desde el sol al planeta barre áreas iguales en tiempos iguales. En consecuencia, el planeta se mueve más despacio en el afelio que en el perihelio. La construcción indicada en el diagrama nos proporciona un resultado semejante. Pero hubieron de añadirse factores correctivos para hacer que la construcción se adaptara más exactamente a los datos recogidos.

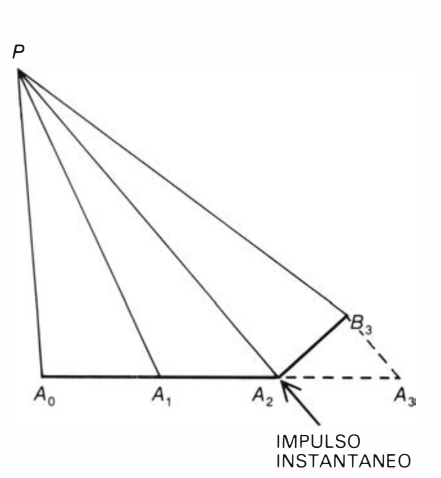
proprietys, and suggest a physicall Reason of this proportion"].

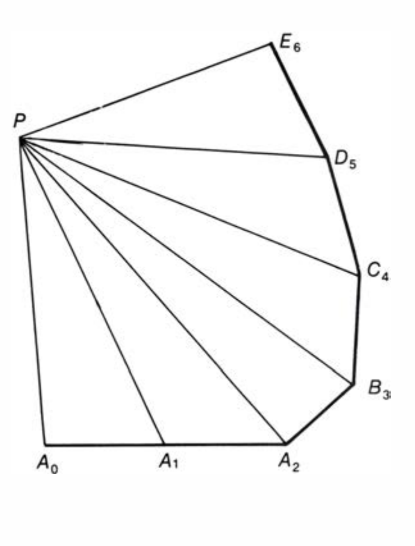
Ciertamente, lo que Newton hizo fue, más o menos, eso. Probó que una elipse satisfacía las condiciones señaladas por Hooke. No obstante, no comunicó el resultado de su prueba a Hooke ni a ningún otro hasta agosto de 1684, cuando le visitó el astrónomo y matemático Edmund Halley. Este vino a ver a Newton para preguntarle "cuál creía él que sería la curva que describirían los planetas, si suponemos que la fuerza de atracción hacia el sol es inversamente proporcional al cuadrado de su distancia a él". El problema había sido muy discutido en la Real Sociedad de Londres. Halley y Christopher Wren eran incapaces de resolverlo, y Hooke nunca presentó una solución, aunque mantenía que la había encontrado.

Cuando Newton oyó la cuestión, respondió inmediatamente: una elipse. Halley le preguntó cómo lo sabía, y Newton replicó: "lo he calculado". Parece que Newton no pudo encontrar los cálculos, pero a instancias de Halley los redactó para la Real Sociedad en el breve tratado *De Motu* (*Sobre el movimiento*). Newton describía en *De Motu* sus investigaciones sobre la dinámica terrestre y celeste, incluyendo sus ideas sobre el movimiento en un espacio vacío y en un medio resistente. El 10 de diciembre de 1684 Newton había concluido necesariamente el *De Motu,* puesto que ese día Halley refirió a la Real Sociedad que Newton le había mostrado recientemente el curioso tratado.

No tenemos documentos precisos sobre el proceso de las ideas de Newton en el intervalo comprendido entre su correspondencia con Hooke y la conclusión del primer borrador del *De Motu*. No obstante, estoy cierto de que fue el método de Hooke para analizar movimientos curvilíneos lo que introdujo a Newton en el recto camino. Aunque no todos los historiadores estén de acuerdo conmigo, yo creo que el tratamiento que hace Newton de la dinámica terrestre y celeste en el *De Motu,* y que desarrolló más ampliamente durante la primavera siguiente en el primer libro de su *Philosophiae Naturalis Principia Mathematica,* refleja el proceso de su pensamiento sobre la dinámica planetaria, inspirado por su correspondencia con Hooke. En un par de manuscritos







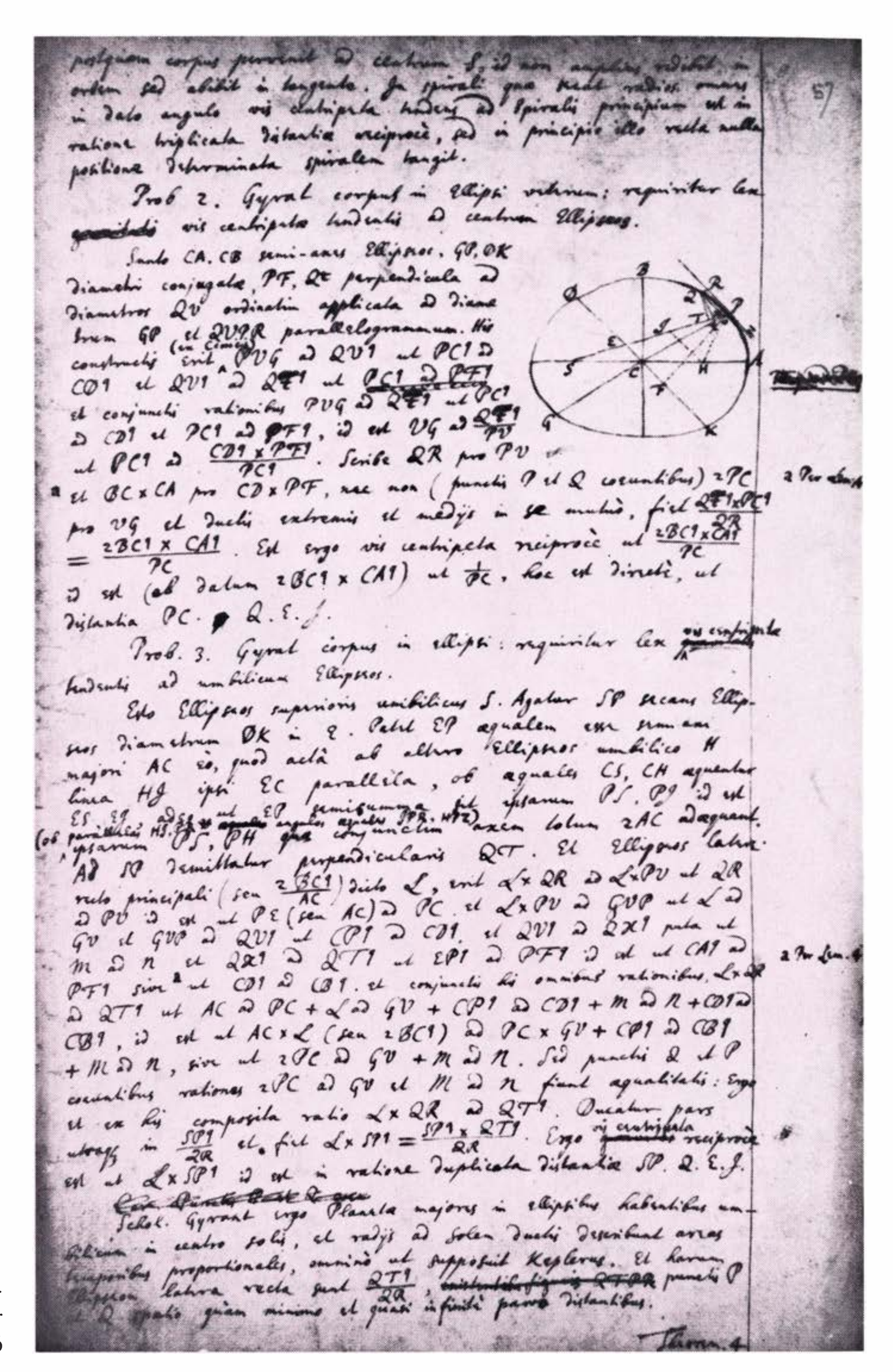
LA FUERZA CENTRIPETA GENERA una trayectoria curva de acuerdo con la ley de las áreas. Esta propiedad de una fuerza centrípeta fue demostrada por Newton en el primer teorema de los *Principia Mathematica* (el título completo del libro es *Philosophiae Naturalis Principia Mathematica*), y en la discusión inicial del breve tratado *De Motu* (*Sobre el Movimiento*). Newton comienza por considerar un cuerpo que se mueve en línea recta y con velocidad constante (*figura de la izquierda*). El cuerpo parte de A_0, y en intervalos sucesivos iguales alcanza primero A_1, después A_2, y así sucesivamente. Se fija un punto P a una distancia h y por encima de esa trayectoria rectilínea. Los triángulos A_0PA_1, A_1PA_2, A_2PA_3, etc., tienen todos la misma área, por tener bases de igual longitud d, y la altura común h. En un segundo estadio del análisis (*figura central*), el cuerpo comienza su movimiento como antes, pero en el momento de llegar a A_2 recibe un impulso instantáneo hacia P. Durante el intervalo siguiente, el cuerpo se moverá también en línea recta, pero ya no alcanzará A_3 sino B_3, que está en la recta A_3B_3 paralela al impulso. Newton muestra que los triángulos A_1PA_2 y A_2PB_3 tienen también la misma área, ya que los triángulos A_2PA_3 y A_2PB_3 tiene la misma base A_2P y alturas iguales. Si el cuerpo recibe un impulso hacia P al fin de cada intervalo, describirá una línea poligonal en torno a P (*figura de la derecha*). Los triángulos que pueden formarse análogamente en este caso resultan también de igual área. En el límite en el que el intervalo entre cada dos impulsos tiende a cero, el cuerpo estará sometido a una fuerza centrípeta dirigida al punto P, y la trayectoria poligonal se convertirá en una curva continua. Pero el área seguirá proporcional al tiempo. Esta prueba reveló la importancia de la ley de las áreas.

autobiográficos, Newton afirma que esa correspondencia o precede o coincide temporalmente con su demostración, publicada primero en el *De Motu,* y después en los *Principia,* de que un objeto que tiene un movimiento inercial y está sometido a una fuerza centrípeta inversamente proporcional al cuadrado de la distancia, se mueve en una órbita elíptica.

Esta demostración es la que resaltó la importancia física de la ley kepleriana de las órbitas elípticas. (Esta ley establece que cada planeta describe una trayectoria elíptica, ocupando el sol uno de los focos de la elipse.) El lector moderno se sorprenderá quizá de que no fuera Kepler, sino Newton, quien pusiera de manifiesto el carácter fundamental de las leyes keplerianas del movimiento planetario. Sin embargo, antes de la publicación de los *Principia,* esas leyes (que incluso eran llamadas hipótesis) no eran tenidas en tanta consideración como vinieron a serlo después.

En particular, la ley kepleriana de las áreas durante el siglo XVII no gozó de ninguna especial categoría. La mayoría de los trabajos astronómicos ni siquiera la mencionan. Por ejemplo, nunca discute esta ley de las áreas, ni alude siquiera a su existencia, la *Astronomia Carolina* de Thomas Streets, de la que Newton transcribió la tercera ley de Kepler. (Según ella, el cubo de la distancia media de un planeta al sol es proporcional al cuadrado del período de su órbita.) La mayoría de los astrónomos del siglo XVII no calculaban las posiciones planetarias mediante la ley de las áreas, sino mediante una construcción basada en un vector que parte del foco no ocupado de la órbita elíptica del planeta y que gira uniformemente. Puesto que los astrónomos empleaban raras veces la ley de las áreas, caer en la cuenta de su importancia exigió una perspicacia extraordinaria por parte de Newton. Fue él quien elevó la ley kepleriana de las áreas a la categoría de que goza hoy día.

El primer teorema de todos en los *Principia* (y la discusión inicial del *De Motu*) desarrolla la importancia dinámica de la ley de las áreas, al probar que el movimiento curvilíneo descrito por esta ley es consecuencia del carácter centrípeto de la fuerza. La prueba, que tiene tres partes, muestra lo bien que había aprendido Newton la técnica de Hooke para descomponer el movimiento curvilíneo en una componente inercial y otra centrípeta.

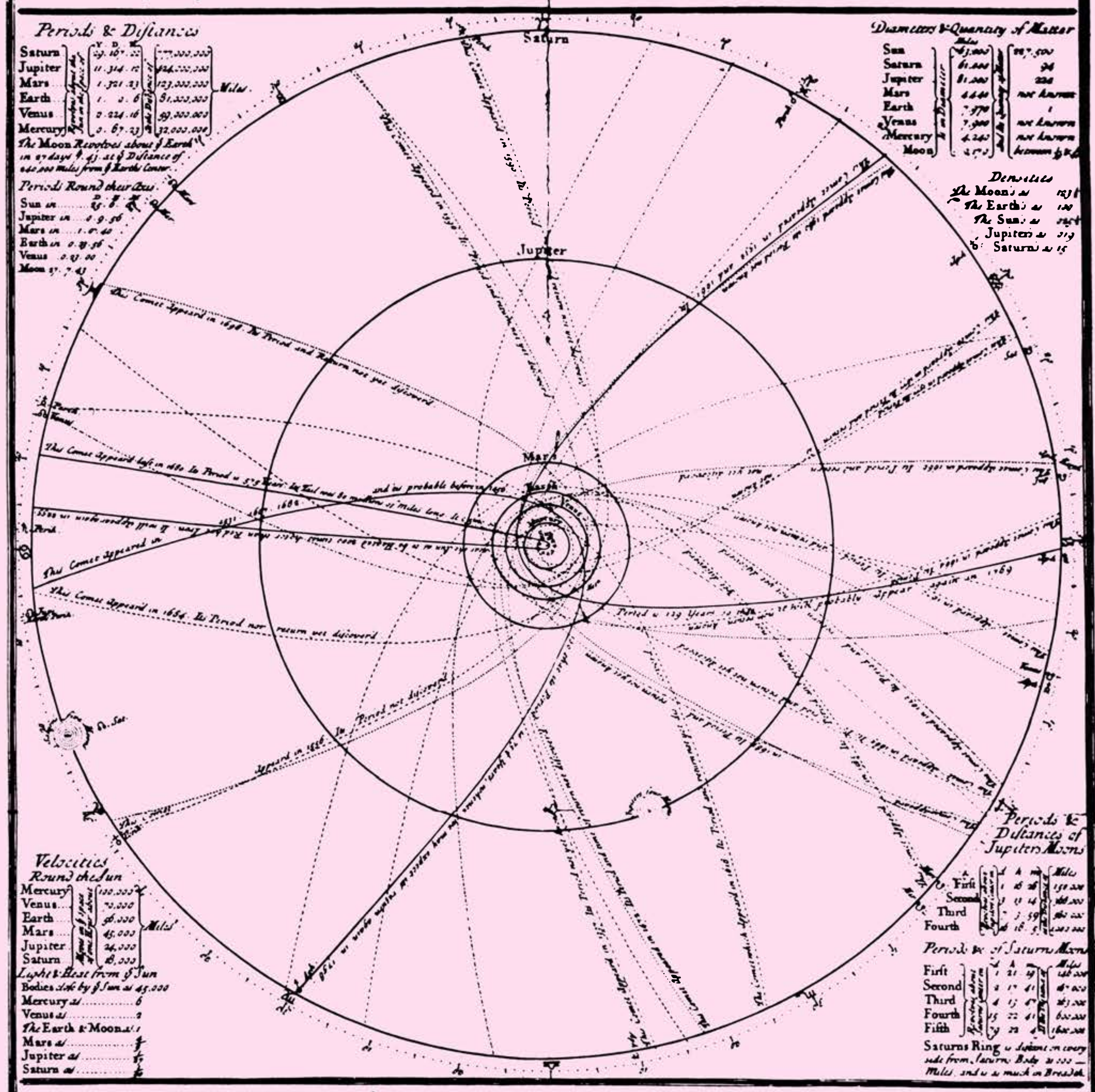


PAGINA de un borrador del "De Motu", escrita de mano del propio Newton, probablemente en noviembre de 1684. En el *De Motu,* discutió la dinámica terrestre y celeste, introduciendo la idea de fuerza centrípeta dirigida al foco de una elipse. La página concluye: "Por consiguiente, los planetas mayores giran en elipses que tienen un foco en el centro del sol, y los radios trazados desde el sol describen áreas proporcionales a los tiempos, enteramente según lo supuso Kepler...". Este escolio es falso, y la naturaleza del error indica que Newton no había elaborado todavía la idea de gravitación universal. Según él mismo advirtió muy pronto, el foco de las órbitas planetarias no es el sol, sino el centro de gravedad común al sol y los planetas. Y además, no sólo atrae el sol a los planetas, sino que también cada planeta atrae al sol.

En la primera parte de la prueba, Newton considera un cuerpo que se mueve en línea recta y con velocidad uniforme. La línea está dividida en intervalos iguales, para indicar que el cuerpo recorre distancias iguales en tiempos iguales. Se fija un punto P a una distancia h de esta línea de movimiento. Los triángulos formados al unir P con cada uno de los intervalos iguales tienen todos la misma área, ya que poseen bases iguales y la altura común h. Con este análisis tan sencillo, Newton pone de manifiesto una relación inespe-

Mr. WHISTON'S SCHEME of the SOLAR SYSTEM EPITOMIS'D. To wch is annex'd

A Tranſlation of part of ye General *Scholium* at ye end of ye ſecond Edition of *Sr. Isaac Newton's* Principia. *Concerning God.*



THIS moſt Elegant Syſtem of the Planets and Comets could not be produced but by and under the Contrivance and Dominion of an Intelligent and Powerful Being. And if the Fixed Stars are the Centers of such other Syſtems, all theſe being Framed by the like Council will be Subject to the Dominion of *One*, eſpecially ſeeing the Light of the Fixed Stars is of the ſame Nature with that of the Sun, and the Light of all theſe Syſtems paſses mutually from one to another. He governs all things, not as ye Soul of the World, but as the Lord of the Univerſe, and becauſe of his Dominion he is wont to be called Lord God παντοκράτωρ (i. e. Universal Emperor) for God is a Relative word, and has Relation to Servants: And the Deity is the Empire of God, n tover his own Body (as is the Opinion of thoſe, who make him the Soul of the World) but over his Servants. The *Supreme God* is a Being Eternal, Infinite Abſolutely Perfect; but a Being however Perfect without Dominion, is not *Lord God*: For we say, *my God, your God, the God of Iſrael*, but we do not ſay, *my Eternal, your Eternal, the Eternal of Iſrael*; we do not ſay *my Infinite, your Infinite, ye Infinite of Iſrael*; we do not ſay *my Perfect, your Perfect, the Perfect of Iſrael*. Theſe Titles have no Relation to Servants. The word God frequently ſignifies Lord, but every Lord is not God. The Empire of a Spiritual being conſtitutes God true Empire conſtitutes True God, Supreme the Supreme, Feigned the Feigned. And from his true Empire it follows that the true God is Living, Intelligent & Powerful, from his other Perfections that he is the Supreme or Supremly Perfect. He is *Eternal & Infinite, Omnipotent* and *Omnipreſent* that is, he endures from Eternity to Eternity, and he is preſent from Infinity to Infinity, he Governs all Things, and Knows all Things which are or which can be known. He is not Eternity or Infinity, but he is Eternal and Infinite, he is not Duration or Space, but he Endures and is Preſent. He endures always and is preſent everywhere and by exiſting always and everywhere, he Conſtitutes Duration and Space, Eternity and Infinity. Whereas every Particle of Space is *always*, and every Individual Moment of Duration is *every where*, certainly the Framer and Lord of the Univerſe ſhall not be (*nunquam nuſquam*) *never no where*. He is Omnipreſent not Virtually only, but alſo Subſtantially, for Power without Subſtance cannot Subſiſt. In him are containd and moved all things (ſo the Antients thought *) but without mutual Paſsion God ſuffers nothing from the Motions of Bodies: Nor do they ſuffer any Reſiſtance from the Omnipreſence of God. It is confeſs'd that the Supreme God exiſts Neceſſarily, and by the ſame Neceſsity he is *always* and *every where*. Whence alſo he is wholy Similar, all Eye, all Ear, all Brain, all Arm, all the Power of Perceiving Underſtanding and Acting. But after a manner not at all Corporeal, after a manner not like that of Men, after a manner wholly to us unknown. As a Blind Man has no notion of Colours, so neither have we any notion of the manner how the moſt Wiſe God perceives and underſtands all things He is wholly deſtitute of all Body and of all Bodily ſhape, and therefore cannot be ſeen, heard, nor touched, nor ought to be Worſhiped under the Repreſentation of any thing Corporeal. We have Ideas of his Attributes, but we know not at all what is the Subſtance of any thing whatever. We ſee only the Figures and Colours of Bodies, we hear only Sounds, we touch only the outward Surfaces, we ſmell only Odours and taſt Taſts, but we know not by any ſence or reflex Act the inward Subſtances, and much leſs have we any Notion of the Subſtance of God: We know him only by his Properties and Attributes and by the moſt Wiſe and Excellent Structure of things, and by Final Cauſes, but we Adore and Worſhip him on account of his Dominion. For God without Dominion, Providence & Final Cauſes, is nothing elſe but Fate and Nature.

* [illegible]

Engrav'd and Sold by John Senex at the Globe in Salisbury Court near Fleetstreet. Where may be had Dr Halley's Scheme of the Total Eclipse of the Sun which will be in 1724. Also his Zodiack containing all the Stars in the Way of the Moon and Planets. Usefull in Astronomical Observations and for finding the Longitude at Sea.

rada entre movimiento inercial y ley de las áreas.

En la segunda parte de la prueba, el móvil se mueve al principio como antes, pero al final del segundo intervalo recibe una fuerza instantánea –un impulso– hacia *P*. Por consiguiente, durante el tercer intervalo el cuerpo ya no se mueve según la recta original, sino según otra recta algo más dirigida hacia *P*. Newton muestra geométricamente que, también ahora, el triángulo formado al unir *P* con el segmento de trayectoria recorrido durante el segundo intervalo tiene la misma área que el formado al unir *P* con el segmento de trayectoria recorrido durante el tercero.

En la tercera parte, el cuerpo recibe un cierto impulso hacia *P*, al final de cada intervalo. Como consecuencia de ello, el cuerpo se desplaza según una trayectoria poligonal que rodea a *P*. Para ella sigue satisfaciéndose la ley de las áreas. En el caso límite en que el intervalo entre los impulsos tiende a cero, el cuerpo estará sometido a una fuerza continua dirigida hacia *P*, y la trayectoria poligonal se convertirá en una curva suave u órbita. De este modo demostró Newton que una fuerza centrípeta genera una trayectoria curva que satisface la ley de las áreas.

El segundo teorema de los *Principia* es el recíproco: Un movimiento a lo largo de una curva descrito por la ley de las áreas implica una fuerza centrípeta. Con estos dos teoremas, Newton demostró que la ley de las áreas es condición necesaria y suficiente del movimiento inercial en un campo de fuerzas centrales.

Los dos teoremas forman parte de una serie de demostraciones que comienzan con la ley de las áreas y concluyen con la prueba de que una órbita elíptica exige una fuerza centrípeta inversamente proporcional al cuadrado de la distancia. Esta serie de demostraciones, presentada tanto en los *Principia* como en el *De Motu,* marca una profunda discontinuidad en la historia de las ciencias exactas. Estas demostraciones introdujeron una dinámica celeste radicalmente nueva, basada en los nuevos conceptos de fuerza, cantidad de movimiento, masa e inercia, y en una medida cuantitativa de la fuerza dinámica enteramente novedosa. El subtítulo de la *Astronomia Nova* de Kepler fijaba el objetivo de crear una "física celeste basada en las causas". Newton alcanzó este objetivo, que Kepler sólo había vislumbrado visionariamente. Ni Galileo ni Descartes habían concebido tal dinámica celeste y la formulación newtoniana dejó muy atrás incluso al gran físico Huygens.

Basándose en el primer borrador del *De Motu,* que Newton escribió probablemente en noviembre de 1684, se demuestra claramente que en ese año no había elaborado todavía el concepto de gravitación universal. El borrador discute la fuerza centrípeta dirigida hacia el foco de una elipse, y concluye con el escolio: "Por consiguiente los planetas mayores giran en elipses que tienen un foco en el centro del sol, y los radios trazados desde el sol describen áreas proporcionales a los tiempos, enteramente según lo supuso Kepler...".

Newton ni probó este escolio, ni continuó creyendo en él por largo tiempo; y, hablando con propiedad, este escolio es falso. Según Newton notó muy pronto, los planetas no se desplazan según la ley de las áreas en órbitas keplerianas simplemente elípticas y con el sol en uno de sus focos. Más bien, el foco está en el centro de gravedad común, ya que no sólo el sol atrae cada planeta, sino que también cada planeta atrae al sol (y además los planetas se atraen unos a otros). Si Newton hubiese formulado ya un principio de gravitación universal, no hubiese propuesto ese escolio erróneo.

Newton cayó muy pronto en la cuenta de que no había probado que el movimiento real de los planetas obedezca exactamente la ley de las órbitas elípticas y la ley de las áreas. Había encontrado únicamente que esas leyes se satisfacen para un sistema de un cuerpo: una sola masa puntual que se mueve en un campo de fuerzas centrales con una componente inicial de movimiento inercial. Reconoció que el sistema de un cuerpo no corresponde al mundo real, sino a una situación artificial que es más fácil de investigar matemáticamente. El sistema de un cuerpo reduce la tierra a una masa puntual y el sol a un centro de fuerzas inmóvil.

Lo que permitió a Newton trascender ese sistema de un cuerpo fue la consideración de las consecuencias de su tercera ley del movimiento: la ley de acción y reacción. Esta ley es quizá la más original de las tres leyes del movimiento (las otras dos son la ley de inercia y la ley de fuerza). Un testimonio de su novedad es que aún hoy es empleada con frecuencia de modo incorrecto por parte de quienes la relacionan, no con una situación de choque o de interacción entre dos cuerpos, sino con una supuesta condición de equilibrio.

El desarrollo de las reflexiones de Newton sobre acción y reacción después de concluir el primer borrador del *De Motu* está expuesto en las secciones iniciales del primer libro de los *Principia.* En la introducción a la sección 11, Newton explica que, de momento, se ha limitado a estudiar una situación que "difícilmente se da en el mundo real", a saber, la de "movimientos de cuerpos atraídos hacia un centro inmóvil". La situación es artificial, ya que "las atracciones comúnmente se dirigen a [otros] cuerpos y –según la tercera ley del movimiento– las acciones de los cuerpos atrayentes y atraídos son siempre mutuas e iguales". En consecuencia, "si hay dos cuerpos, ni el cuerpo atrayente ni el cuerpo atraído pueden estar en reposo". Más bien, "ambos cuerpos (por el cuarto corolario de las leyes) dan vueltas en torno a un centro común, como si lo hicieran por una atracción mutua".

Newton había visto que si el sol tira de la tierra, la tierra ha de tirar también del sol, con una fuerza de igual magnitud. En este sistema de dos cuerpos, la tierra no se desplaza en una órbita simple en torno al sol. Más bien, tanto el sol como la tierra se mueven en torno a su centro de gravedad común. Otra consecuencia de la tercera ley del movimiento es que cada planeta es un centro de fuerza atractiva, del mismo modo que es un cuerpo atraído. Por consiguiente, un planeta no sólo atrae a, y es atraído por, el sol, sino que también atrae a, y es atraído por, cada uno de los otros planetas. Aquí Newton ha dado ya el paso decisivo desde un sistema interactivo de dos cuerpos hasta un sistema interactivo de muchos cuerpos.

En diciembre de 1684 Newton

SISTEMA NEWTONIANO DEL MUNDO, dibujado por William Whiston, que sucedió a Newton en su cátedra "Lucasiana" de la Universidad de Cambridge. El diagrama está tomado del pliego de Whiston "Esquema condensado del sistema solar", publicado en 1724. Los planetas y los satélites de Júpiter y Saturno aparecen girando en torno al sol, bajo la acción de la gravitación universal. Con notable perfección incluyó Whiston también las órbitas de los cometas. Newton había demostrado que estas órbitas son elipses o parábolas, en las que un radio vector del sol al cometa barre áreas iguales en tiempos iguales. Debajo del diagrama se reproduce, en traducción inglesa de Whiston, un fragmento final del Escolio General de los *Principia* (correspondiente a la segunda edición de 1713). En él escribió Newton que "Este elegantísimo sistema de Planetas y Cometas no pudo ser producido más que por y bajo la Habilidad y el Dominio de un Ser Inteligente y Poderoso". (Ilustración cedida por Goldsmith Company's Library.)

concluyó un borrador revisado del *De Motu* que describe el movimiento planetario en el contexto de un sistema interactivo de muchos cuerpos. A diferencia del primer borrador, el revisado concluye que "los planetas ni se mueven exactamente en elipses, ni giran dos veces según la misma órbita". Esta conclusión le condujo a Newton al resultado siguiente: "hay tantas órbitas para un planeta cuantas revoluciones según ocurre en el movimiento de la luna, y la órbita de cada planeta depende del movimiento combinado de todos los planetas, por no mencionar las acciones de todos ellos unos sobre otros". Y a continuación: "considerar simultáneamente todas estas causas de movimiento y definir estos movimientos mediante leyes exactas que permitan un cálculo apropiado, si no me equivoco, excede la capacidad de todo el entendimiento humano".

No se conservan documentos que indiquen cómo Newton, durante el mes o mes y pico transcurrido entre la redacción del primer borrador del *De Motu* y su revisión, llegó a concebir que los planetas actúan gravitacionalmente unos sobre otros. No obstante, el pasaje antes citado expresa esa concepción en un lenguaje inequívoco: "*eorum omnium actiones in se invicem*" ("las acciones de todos ellos unos sobre otros"). Una consecuencia de esta mutua atracción gravitacional es que ninguna de las tres leyes de Kepler es estrictamente verdadera en el mundo de la física, sino que sólo son verdaderas en un modelo matemático en el que masas puntuales que no interaccionan entre sí giran en torno a un centro de fuerzas matemático, o un cuerpo atractivo estacionario. La distinción que traza Newton entre el reino de las matemáticas, donde las leyes de Kepler son verdaderas leyes, y el reino de la física, donde son sólo "hipótesis" o aproximaciones, es uno de los caracteres revolucionarios de la dinámica celeste newtoniana.

He supuesto que la tercera ley del movimiento fue el factor clave en el razonamiento que condujo a Newton a sugerir perturbaciones gravitacionales mutuas en las órbitas planetarias. No tenemos prueba directa de esta suposición, ya que no existen documentos que contengan una versión previa de su frase "las acciones de todos ellos unos sobre otros". Sin embargo, tenemos una prueba indirecta muy fuerte. En el verano de 1685, pocos meses después de revisar el *De Motu,* Newton estaba ciertamente a punto de terminar el primer borrador de los *Principia.* En la versión primitiva de lo que fue a convertirse en un segundo libro, "El sistema del mundo", Newton refleja paso a paso el proceso que le condujo a la idea de interacciones gravitacionales de los planetas. En este proceso, la tercera ley del movimiento desempeña el papel central, y yo no veo motivo para creer que no es el mismo proceso que le condujo de hecho a esa idea pocos meses antes, al revisar el *De Motu.*

He aquí dos pasajes del primer borrador de "El sistema del mundo" (redactado en latín), que manifiestan el papel crucial que desempeñó en esto la tercera ley del movimiento:

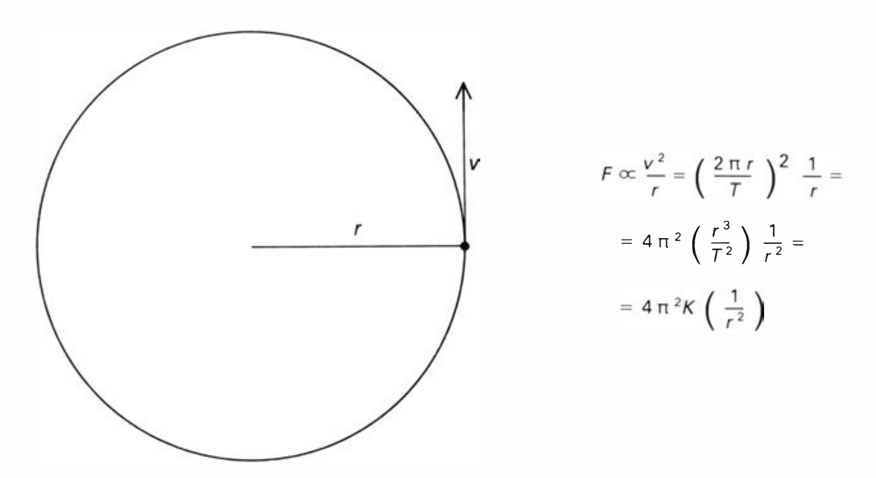
"20. *La concordancia entre las analogías.*

"Y puesto que la acción de la fuerza centrípeta sobre el cuerpo atraído, para distancias iguales, es proporcional a la materia de este cuerpo, es razonable también que sea asimismo proporcional a la materia del cuerpo que atrae. Puesto que la acción es mutua, hace que los cuerpos se acerquen uno a otro por un comportamiento mutuo (según la ley 3), y consiguientemente debe ser semejante a sí misma en ambos cuerpos. Podemos considerar un cuerpo como atrayente y otro como atraído, pero esta distinción es más matemática que natural. La atracción es en realidad la de cada uno de los cuerpos respecto al otro, y así es del mismo tipo en cada uno de los cuerpos."

"21. *Y su identificación.*

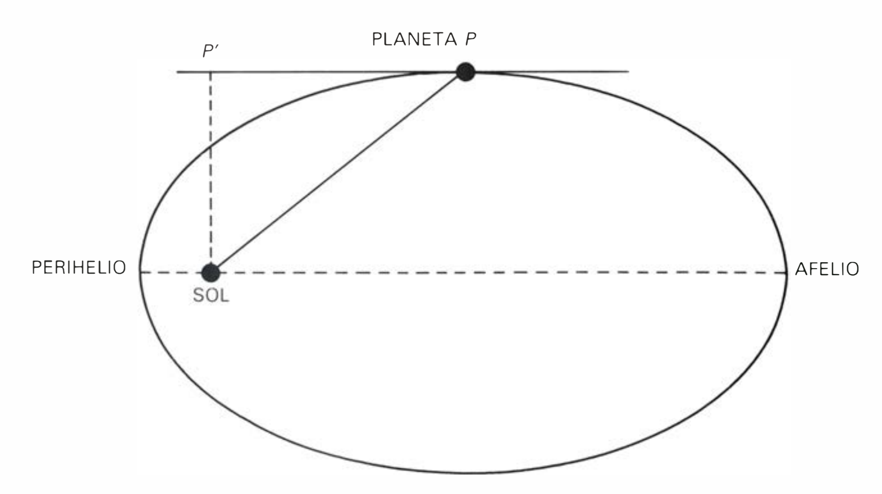
"Y de ahí que la fuerza atractiva se encuentre en ambos cuerpos. El sol atrae a Júpiter y a los otros planetas, Júpiter atrae sus satélites y, análogamente, lo hacen los satélites entre sí y sobre Júpiter, y los planetas todos entre sí. Y aunque para una pareja de planetas las acciones de cada uno de ellos en el otro pueden distinguirse una de otra y pueden considerarse como dos acciones por las que cada planeta atrae al otro, sin embargo en cuanto son acciones entre los dos mismos cuerpos no son dos operaciones sino una sola entre dos términos. Dos cuerpos pueden tirar uno de otro por la contracción de una cuerda entre ambos. La causa de esta acción es doble, a saber, la disposición de cada uno de los dos cuerpos; la acción es asimismo doble, en cuanto actúa sobre dos cuerpos; pero en cuanto actúa entre dos cuerpos es una única acción. No hay, por ejemplo, una operación por la que el sol atrae a Júpiter y otra por la que Júpiter atrae al sol, sino una única operación por la que el sol y Júpiter tienden a acercarse uno a otro. Por la acción por la que el sol atrae a Júpiter, Júpiter y el sol tienden a acercarse uno a otro (según la ley 3), y por la acción por la que Júpiter atrae al sol, Júpiter y el sol tienden también a acercarse uno a otro. Sin embargo, el sol no es atraído hacia Júpiter por una acción doble, ni Júpiter es atraído hacia el sol por una acción doble, sino que hay una única acción entre ellos por la que ambos se acercan entre sí."

A continuación, concluye Newton que "según esta ley todos los cuerpos han de atraerse entre sí". Presenta con orgullo su conclusión y explica por qué la magnitud de la fuerza atractiva es tan pequeña que resulta inobservable. "Sólo es posible observar estas fuerzas en los cuerpos enormes de los planetas".



NATURALEZA CUADRATICO-INVERSA DE LA FUERZA CENTRIPETA para el caso de órbitas circulares. Puede deducirse fácilmente de la tercera ley de Kepler sobre el movimiento planetario y de la ley de la fuerza centrípeta. Según la tercera ley de Kepler, r^3/T^2 es una constante K, si llamamos r al radio de la órbita planetaria y, T, al período de la órbita. La ley de la fuerza centrípeta establece que, para una trayectoria circular, la fuerza centrípeta, F, es proporcional a v^2/r, siendo v la velocidad del planeta. En el período T el planeta describe una órbita completa, de longitud $2\pi r$, por lo que su velocidad v vale $2\pi r/T$. Esto basta para deducir que la fuerza centrípeta es inversamente proporcional al cuadrado del radio r^2.

En el libro tercero de los *Principia,* que estudia también el sistema del mundo, aunque de modo algo más matemático, Newton trata el tema de la gravitación prácticamente de la misma manera. Primero, en el llamado "ensayo lunar" ("moon test"), extiende a la luna la fuerza del peso, o gravedad terrestre, y demuestra que esta fuerza varía en razón inversa del cuadrado de la distancia. Identifica luego esa fuerza terrestre con la fuerza que ejerce el sol sobre los planetas y la que ejerce un planeta sobre sus satélites. A todas estas fuerzas las llama ahora gravedad. Con ayuda de la tercera ley del movimiento, transforma la idea de una fuerza del sol sobre los planetas en la de una fuerza mutua entre el sol y los planetas. Análogamente, transforma la idea de una fuerza de los planetas sobre sus satélites en la de una fuerza mutua entre los planetas y sus satélites, y entre estos últimos entre sí. La transformación final es la noción de que todos los cuerpos interaccionan gravitacionalmente.



VELOCIDAD ORBITAL DE UN PLANETA. Tal velocidad no es inversamente proporcional a la distancia ordinaria *SP*, entre el sol y el planeta, sino que, debido a la ley de las áreas, es inversamente proporcional a la "distancia perpendicular", es decir, a la distancia *SP'* entre el sol y la tangente *PP'* a la órbita planetaria en el punto *P*. Sólo para dos puntos de la órbita, para el perihelio y para el afelio, la distancia ordinaria y la distancia perpendicular coinciden. (Ilustración realizada por Alan D. Iselin.)

Mi análisis de los estadios del pensamiento de Newton no ha de tomarse como una minimización del poder extraordinario de su genio creativo; más bien debería hacer plausible tal genio. El análisis muestra lo fecundo del modo newtoniano de pensar en física, en el que las matemáticas se aplican al mundo externo tal como éste se manifiesta por la experiencia y la observación crítica. Este modo de pensar, que yo llamo el "estilo newtoniano", está bien expresado en el título original de la gran obra de Newton: *Principios Matemáticos de Filosofía Natural.*

El estilo newtoniano consiste en un repetido ir y venir entre un modelo matemático y la realidad física. En el desarrollo de las ideas de Newton sobre la gravedad y en su presentación de otras ideas en los *Principia,* partió de un modelo matemático que representa una simplificación de la naturaleza: una masa puntual moviéndose en torno a un centro de fuerzas. Al no suponer que el modelo sea una representación exacta del universo físico, Newton gozaba de libertad para explorar las propiedades y efectos de una fuerza atractiva de carácter matemático, a pesar de que, para él, el concepto de una fuerza que tira "actuando a distancia" era algo repugnante e inadmisible en el campo de una buena física. A continuación comparó las consecuencias de su modelo matemático con los principios y leyes observadas en el mundo externo, como la ley kepleriana de las áreas y la de las órbitas elípticas. Donde el modelo matemático resultaba insuficiente, Newton lo modificó. Hizo que el centro de fuerza no fuera una entidad matemática, sino una masa puntual. Digo una masa puntual, más que un cuerpo físico, porque no había considerado todavía propiedades físicas tales como tamaño, forma y masa.

A partir del modelo matemático modificado, Newton concluyó que un conjunto de masas puntuales se atraían mutuamente y, mutuamente, perturbaban sus órbitas. De nuevo comparó el modelo con el mundo físico. De todos los planetas, Júpiter y Saturno son los más masivos, y por ello pudo observar perturbaciones en sus movimientos orbitales. Con la ayuda de John Flamsteed, Newton encontró que el movimiento orbital de Saturno se perturba cuando los dos planetas están próximos. El proceso de comparar una y otra vez el modelo matemático con la realidad y luego modificarlo convenientemente condujo a Newton a considerar los planetas como cuerpos físicos con formas y tamaños definidos.

Después de haber modificado su modelo, Newton lo aplicó muchas veces al sistema del mundo. Aseguró que la fuerza de atracción, que había derivado matemáticamente, era la gravitación universal. Encontró que la luna se movía como si fuera atraída hacia la tierra con una fuerza que era 1/3600 de la fuerza de la gravedad con la que la tierra tiraba de los objetos colocados en su superficie. Puesto que la luna está 60 veces más lejos del centro de la tierra que los objetos colocados en la superficie de ésta, el factor 1/3600 concuerda con la deducción de que la gravedad de la tierra se extiende a la luna y disminuye con el cuadrado de la distancia.

La ley de gravitación universal explica por qué los planetas obedecen aproximadamente las leyes de Kepler y por qué se apartan de ellas en la forma en que lo hacen. Da razón de por qué (en ausencia de rozamientos) todos los cuerpos caen con el mismo aceleramiento en un lugar cualquiera de la tierra, y por qué ese aceleramiento varía con la altura y con la latitud. La ley de gravitación explica también los movimientos regulares e irregulares de la luna, proporciona una base física para entender y predecir las mareas y muestra cómo el ritmo de precesión de la tierra, observado desde mucho antes pero nunca explicado, es efecto de que la luna tira de la convexidad ecuatorial de la tierra. Puesto que la fuerza matemática de atracción va bien para explicar y predecir los fenómenos observados del mundo, Newton decidió que esa fuerza tenía que "existir de verdad", aun cuando la filosofía tradicional a la que él se adhería no permitía ni podía permitir que tal fuerza formase parte de un sistema de la naturaleza. Y por eso exigía una investigación sobre cuál podía ser el origen de los efectos de la gravitación universal.

Aunque Newton pensó a veces que la gravitación universal podía originarse por impulsos de una corriente de partículas de éter que bombardea un objeto, o por modificaciones de un éter que todo lo penetra, no quiso adelantar ninguna de esas dos ideas en los *Principia,* fiel a su consigna de "no imaginar hipótesis" como explicaciones físicas. El estilo newtoniano le había llevado a

un concepto matemático de fuerza universal, y este mismo estilo le llevó a aplicar su resultado matemático al mundo físico, aun cuando ese tipo de fuerza no resultara muy creíble.

Algunos contemporáneos de Newton quedaron tan turbados por la idea de una fuerza atractiva que actúa a distancia, que no pudieron ponerse a explorar sus propiedades, y les resultó muy difícil aceptar la física newtoniana. No pudieron seguir a Newton cuando afirmaba que no había sido capaz de explicar cómo actúa la gravitación, aunque "es suficiente que la gravitación exista realmente y pueda explicar los fenómenos de los cielos y las mareas". Los que aceptaban el estilo newtoniano se adiestraron en el manejo de la ley de gravitación universal, mostraron cómo explicaba muchos otros fenómenos físicos y exigieron que se buscase una explicación de cómo podía transmitirse esa fuerza a enormes distancias a través de espacios aparentemente vacíos. El estilo newtoniano le permitió a Newton estudiar la gravitación universal sin inhibiciones prematuras que habrían bloqueado su gran descubrimiento. El biólogo del siglo XVIII George Louis Leclerc de Buffon escribió una vez que el estilo de un hombre no puede distinguirse del hombre mismo. En el caso de Newton, su mayor descubrimiento no puede separarse de su estilo.

La correspondencia entre Hooke y Newton muestra claramente que Hooke enseñó a Newton cómo analizar movimientos curvilíneos. Hooke posteriormente sostuvo una pretensión mucho más exigente: que se le reconociera el mérito de haber sugerido a Newton la ley de gravitación universal inversamente proporcional al cuadrado de la distancia. Muchos historiadores han orquestado el punto de vista de Hooke.

Esa pretensión, sin embargo, no puede sostenerse. Hooke había sugerido simplemente que los planetas están sometidos a una fuerza cuadrático-inversa dirigida hacia el sol. La gravitación universal es mucho más que una fuerza dirigida hacia el sol. Implica también un efecto de los planetas sobre el sol. Y además, se aplica a todos los objetos del universo. La ley de gravitación universal no es una simple relación cuadrático-inversa; es también una relación matemática entre las masas de los cuerpos que se atraen. Fue necesaria una enorme intuición para saltar de una fuerza cuadrático-inversa dirigida hacia el sol a la gravitación universal. Y fue necesario el genio de Newton para inventar el concepto moderno de masa.

Newton no se sintió deudor de Hooke ni siquiera por haberle sugerido que la fuerza centrípeta era inversamente proporcional al cuadrado de la distancia. Huygens había publicado en 1673 un suplemento de su libro sobre el reloj de péndulo en el que afirmaba que, para el movimiento circular, la medida de la fuerza centrífuga venía dada por v^2/r, donde v era la velocidad del cuerpo que gira y r era el radio de giro. Newton había descubierto independientemente la misma relación por los años 1660. Puesto que la diferencia matemática entre una fuerza centrífuga y una centrípeta es sólo cuestión de su sentido, la relación v^2/r vale también para la fuerza centrípeta. De esta relación y de la tercera ley de Kepler se obtiene, por simple álgebra, que la fuerza centrípeta varía en proporción inversa al cuadrado de la distancia. Tras publicarse el libro de Huygens, cualquiera que tuviese un conocimiento rudimentario de álgebra podía haber encontrado una fuerza centrípeta cuadrático-inversa para el caso de una órbita circular. Según eso, Newton no vio necesidad de reconocer el mérito de la afirmación de Hooke sobre una ley cuadrático-inversa.

Hooke y Newton sabían que hallar una ley cuadrático-inversa para órbitas circulares no era lo mismo que mostrar que la ley valía también para órbitas elípticas en las que el movimiento obedece la ley kepleriana de las áreas. La tarea que llevó a cabo Newton consistió en demostrar que una ley cuadrático-inversa para una fuerza centrípeta correspondía a un movimiento de giro según las leyes keplerianas de las órbitas elípticas y de las áreas. Al discutir este punto en su carta del 6 de enero de 1680, Hooke cometió un error fundamental, fallo que sirvió para convencer a Newton de que Hooke no acababa de entender de qué estaba hablando. Hooke decía que si la atracción variaba en razón inversa del cuadrado de la distancia, la velocidad orbital de un planeta sería, "como Kepler supone, *inversamente* proporcional a la *distancia*" ["as Kepler supposes Reciprocall to the Distance"]. Ahora bien, bajo las condiciones que suponía Hooke la velocidad orbital no era inversamente proporcional a la simple distancia del sol al planeta, fuera de los puntos extremos de la órbita: el perihelio y el afelio. En vista del error de Hooke, Newton no estaba para reconocimientos por haberle sugerido la naturaleza cuadrático-inversa de la fuerza centrípeta.

En 1717 Newton quiso asegurar su propia prioridad en el descubrimiento de la ley cuadrático-inversa de gravitación, y para ello montó una escenografía donde realizar el famoso ensayo lunar, no al escribir los *Principia,* sino dos décadas antes, en los años 1660. Sin embargo, los documentos de los años 1660 muestran que Newton no estaba comparando entonces la caída de la luna a lo largo de su órbita con la caída de los objetos terrestres, sino que estaba comparando el "comportamiento centrífugo" de la luna en su giro con el "comportamiento centrífugo" de un cuerpo colocado en la superficie de la tierra que girase según el movimiento diurno de la tierra. Calculó que, para órbitas planetarias circulares, el "comportamiento centrífugo" sería inversamente proporcional a la distancia del sol al planeta, pero no sacó ninguna conclusión física de este cálculo.

Newton nunca publicó su montaje ficticio del antiguo ensayo lunar. Lo incluyó en el borrador manuscrito de una carta al escritor francés Pierre Des Maizeaux, pero después lo tachó. Newton también hizo circular el episodio familiar de que la caída de una manzana le sumió en una cadena de reflexiones que condujeron al descubrimiento de la gravitación universal. Es de suponer que esta invención formaba también parte de su campaña por retrotraer el descubrimiento de la gravitación a un tiempo 20 años anterior a los *Principia.*

Las raíces reales del descubrimiento no pueden anteceder en nada a diciembre de 1684, cuando Newton cayó en la cuenta por primera vez de que, si el sol atrae la tierra, la tierra ha de atraer al sol con una fuerza de igual magnitud. En 1685 superó su habitual resistencia a redactar sus descubrimientos y comenzó a hacer el borrador de los *Principia* para su publicación por la Real Sociedad de Londres. Quizá su complacencia en presentar esta obra a la consideración pública (y con ello arriesgarse a un posible rechazo) estaba motivada en primer lugar por su trascendental descubrimiento de las perturbaciones interplanetarias que se seguían de su atrevida concepción de la gravitación universal. Tenía entre las manos el fundamento de un nuevo sistema de filosofía natural que podía desarrollarse sobre principios matemáticos. En resumen, apenas tuvo Newton algo que decir realmente importante sobre la dinámica celeste, estaba deseoso y aun impaciente por presentarlo al mundo.

Temas metamágicos

Por algoritmo o por tanteo, cubistas y cubomaestros tenaces no paran de darle vueltas al Cubo Mágico

Douglas R. Hofstadter

Cubitis magiquia, *f.* Grave trastorno mental, acompañado casi siempre de sensaciones pruriginosas en las yemas de los dedos; sólo puede aliviarse mediante contacto prolongado con un cubo multicolor originario de Hungría y Japón. El síndrome suele persistir durante meses. Sumamente contagiosa.

Lo que esta parrafada pseudomédica ha olvidado mencionar es que el contacto con el cubo multicolor no sólo puede calmar la desazón, sino, sobre todo, generarla. Además, tampoco señala que tal aflicción puede ser sumamente deleitosa. Creo saber de qué hablo: la padecí durante todo el año pasado, y aún son perceptibles algunos de los síntomas.

Bűvös Kocka –el Cubo Mágico, conocido también por "cubo de Rubik"– ha tomado por asalto, a un mismo tiempo, al mundillo de los pasatiempos y rompecabezas, al mundo de las ciencias de cómputo e inteligencia artificial y al mundo matemático. Pocas veces ha podido un rompecabezas inflamar así la imaginación de tantas personas desde la aparición del famoso panel de 15 piezas creado por Sam Loyd, culpable del ramalazo de locura contagiosa que afligió al mundo civilizado durante el siglo pasado, y que todavía hoy es uno de los pasatiempos más conocidos. El rompecabezas de Loyd y el Cubo Mágico pertenecen a un mismo linaje espiritual. El primero es un problema plano, que consiste en situar correlativamente las piezas de un cuadro de 4 por 4; el otro es tridimensional, y pide reordenar por colores las piezas desordenadas de un cubo de colores de 3 por 3 por 3. Ambos problemas exigen de quien se proponga resolverlos estar dispuesto a deshacer una y otra vez progresos aparentemente muy valiosos; no hay forma de lograr el objetivo sin tener que destruir temporalmente parte de lo conseguido hasta entonces. Esta lección ya resulta dura con las quince piezas de Loyd; con el Cubo Mágico es casi intolerable. Y ambos rompecabezas tienen la pérfida propiedad de que tanto un manazas bien intencionado como un listillo con mala idea puede desmontarlo y volverlo a montar en posiciones de aspecto inofensivo, desde las que es imposible por completo alcanzar la organización deseada, provocando así en el aspirante a maestro cubero la frustración más angustiosa.

El Cubo Mágico es mucho más que un simple rompecabezas. Es un ingenioso invento mecánico, es un pasatiempo, un instrumento de aprendizaje, una fuente de metáforas, una inspiración. Hoy nos parece objeto familiar, mas hizo falta largo tiempo para descubrirlo. Sin duda el tiempo de la cosecha había llegado ya, pues la idea germinó y se desarrolló casi paralelamente en Hungría y Japón y, tal vez, también en otros lugares. Hace poco se publicó un informe que daba cuenta de que un funcionario francés recordaba haber encontrado un cubo así, construido con piezas de madera, en Estambul, en 1920, y de nuevo, más tarde, en Marsella, en 1935. De no tener más amplia confirmación estos informes no nos parecen muy de fiar. En todo caso, el trabajo de Rubik quedó listo en 1975, y su patente húngara data de esa fecha. Por otra parte, y con total independencia, Terutoshi Ishige, ingeniero autodidacta y propietario de una pequeña fundición cercana a Tokio, se presentó un año más tarde con un diseño muy parecido al de Rubik, solicitando patente japonesa en 1976. También Ishige merece que le sea reconocida su extraordinaria intuición.

Pero, ¿quién es Rubik? Ernő Rubik es profesor en la Escuela de Artes Comerciales de Budapest, donde enseña arquitectura y diseño. Buscando aguzar la capacidad de sus alumnos para visualizar objetos tridimensionales, dio con la idea de construir un cubo de $3 \times 3 \times 3$, cada una de cuyas seis caras (de 3×3) pudiese girar alrededor de su centro, y ello sin que el cubo, en su conjunto, se descompusiera en piezas. Inicialmente cada cara estaría pintada de un solo color, pero al girar reiteradamente las diversas caras, los colores quedarían terriblemente revueltos. Sus alumnos tendrían entonces que imaginar cómo desenredarlos.

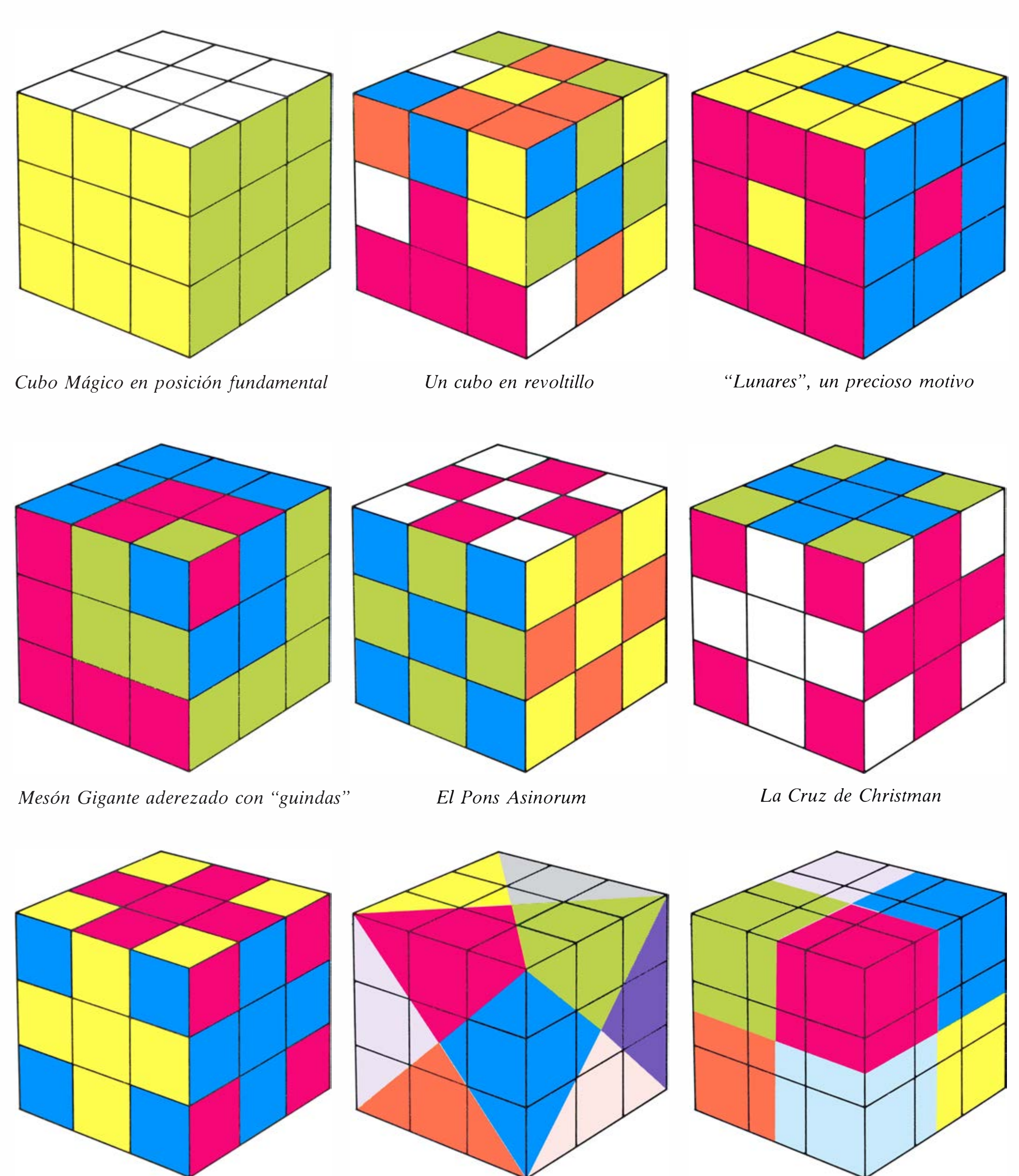
La primera vez que me hablaron del cubo, por teléfono, me pareció ser un imposible físico. La lógica más aplastante hacía ver que el cubo tendría necesariamente que desmenuzarse en sus "cubitos" componentes. Pues fijémonos en uno de los vértices. ¿Cómo se mantiene en su lugar el cubito que lo ocupa? Dado que pertenece a tres caras, haciéndolas girar vemos que el cubito en cuestión ha de poderse separar de sus tres vecinos situados sobre las aristas. ¿Por qué permanece entonces en su sitio? Hay quienes piensan que es gracias a imanes, a bandas elásticas o a complejos sistemas de cables flexibles que atraviesan el interior del cubo; empero, su diseño es extraordinariamente sencillo y no se vale de ninguno de estos elementos.

La verdad es que el Cubo Mágico sí puede desmontarse en pocos instantes [*véase la ilustración inferior de la página 125*], dejando a la vista una estructura interior tan simple que es necesario detenerse a pensar cómo tan sencillo mecanismo puede conseguir tanto como logra. Para comprender cómo se aguanta de una pieza, observemos primero que hay en él tres tipos de cubitos: seis que son centros de caras, 12 que son centros de aristas y ocho que son vértices. Los cubitos centrales tienen, en realidad, tan sólo una cara o faceta, los centros de arista tienen dos y los situados en los vértices, tres. Además, los cubitos centrales no son verdaderamente cúbicos, sino simples fachadas sustentadas por ejes mutuamente ortogonales que parten de una cruceta séxtuple situada en el centro geométrico del cubo. Los restantes cubitos son casi perfectamente cúbicos, con la salvedad de que cada uno tiene una patita roma que apunta hacia el centro del cubo, y también ciertas ranuras curvadas que dan al interior.

El truco básico consiste en que los cubitos se mantienen unos a otros gracias a sus patas, pero sin que ninguno esté ligado a ningún otro. Los cubitos de arista sujetan las patas de los situados en los vértices, y éstos las de aquéllos. Los cubitos centrales de las caras son las piezas maestras de la estructura. Al hacer girar una de las capas –la superior, pongamos por caso– por sí misma, se mantiene formando un todo horizontalmente, mientras que en sentido vertical la mantiene la capa ecuatorial situada bajo ella. La capa ecuatorial

tiene una pista circular excavada en ella (cuyos tramos son las ranuras curvadas de sus cubículos) que guía el movimiento de las patitas de la capa superior y contribuye a mantenerla en su sitio y de una pieza.

En su magistral tratamiento del Cubo Mágico, titulado "Notes on Rubik's 'Magic Cube'", David Singmaster, profesor de ciencias matemáticas y de cómputo en el Politécnico de South Bank, en Londres, define como "problema mecánico fundamental" el de idear cómo puede ser construido el Cubo. Me pregunto a veces si el problema de visualización que Rubik se proponía presentar a sus alumnos sería realmente el de reorganizar las caras (que Singmaster llama "problema matemático fundamental") o si se trataría del problema mecánico. Tengo la impresión de que, entre ambos, el segundo es el más difícil. Yo mismo habré dedicado alrededor de 50 horas de trabajo, repartidas en varios meses, antes de lograr resolver el problema de reordenación; no logré dar con la solución del problema



Cubo Mágico en posición fundamental

Un cubo en revoltillo

"Lunares", un precioso motivo

Mesón Gigante aderezado con "guindas"

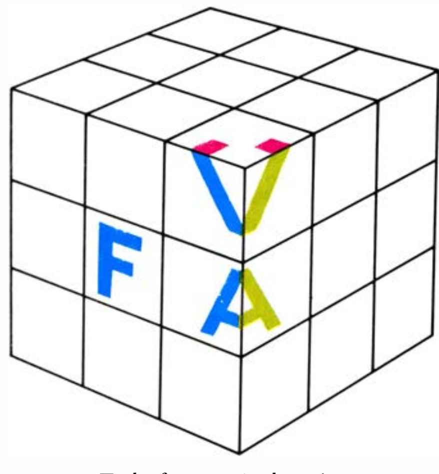
El Pons Asinorum

La Cruz de Christman

La configuración Cruz de Plummer

Coloreado alternante con 12 colores

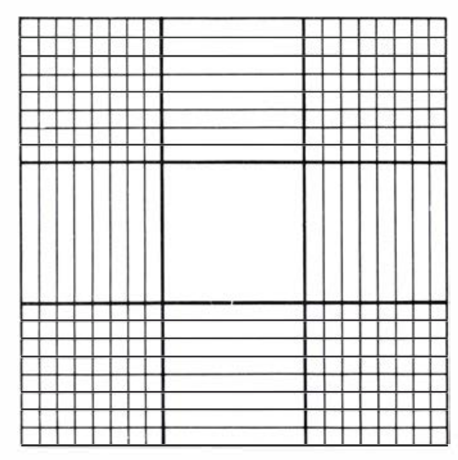
Un coloreado alternante con ocho tonos



F *de faceta,* **A** *de arista*
y **V** *de vértice*

mecánico hasta que no vi el cubo desarmado. Así a ojo, Singmaster estima que las personas que finalmente logran resolver el problema de reorganización (sin ayudas ni indicaciones) invierten en promedio un par de semanas de esfuerzo sistemático. Evidentemente, a quienes lo han conseguido les resulta difícil decir exactamente cuánto tiempo invirtieron (¿cómo distinguir juego de trabajo en casos así?), pero podemos afirmar prudentemente que, de estar condenados a resolver a toda costa el problema de reordenación, harían falta entre cinco horas y un año completo de esfuerzos. Espero que saberlo tranquilice al lector.

Al principio, muchas personas no alcanzan a darse cuenta de que devolver el cubo a la posición de partida (estado en que cada cara es de un único color) es tan difícil que, por lo general, es preciso descubrir un algoritmo capaz de lograrlo a partir de cualquier estado, por entremezclados que estén los colores. Nadie es capaz de reorganizar el cubo por puro tanteo. Quienquiera que haya sido capaz de devolver el cubo a su prístino estado inicial habrá tenido que construirse una pequeña teoría científica.



Cualquiera de las caras del grupo sección central

Una advertencia: suele ocurrir que las soluciones propuestas para el problema mecánico carezcan de claridad, y tengan, o demasiado detalle, o demasiado poco. Ciertamente, dar con un mecanismo que posea la libertad de giro y torsión que tiene el Cubo Mágico es duro reto, pero quizá no lo sea menos la descripción del mecanismo mediante explicaciones y diagramas que puedan ser rápidamente comprendidos por otras personas. Por idéntica razón, para dar algoritmos capaces de reordenar el cubo es imprescindible disponer de notaciones claras y eficaces. El propio Singmaster ha creado una excelente, hoy considerada estándar, que presentaré más adelante. Un segundo aviso: yo no soy "maestro cubero" (que por definición son quienes han contribuido a crear la difícil ciencia de la cubología), sino tan sólo un "cubista", un aficionado pasmado por el cubo y por los virtuosos que han alcanzado a dominarlo a fondo. Por consiguiente, no soy destinatario idóneo para juzgar sobre soluciones nuevas, ni del problema mecánico ni del de reordenación. A los lectores que crean haber atinado con ideas nuevas les recomiendo se pongan en contacto con Singmaster, quien periódicamente va poniendo al día su librito. La dirección de Singmaster es: Department of Mathematical Sciences and Computing, Polytechnic of the South Bank, London SE1 OAA, Inglaterra.

Confío en que tras este aperitivo los lectores tengan la boca tan hecha agua que la posesión de un Cubo sea considerada urgente. Es probable que mucha gente que ya lo ha adquirido no barrunte tan siquiera la profunda dificultad del "problema matemático fundamental". Harán girar despreocupadamente cuatro o cinco caras; de pronto se encontrarán perdidos sin remedio. Tal vez comenzarán entonces a voltear frenéticamente las caras, una tras otra, de todas las formas imaginables, al tiempo que en sus mentes irán percibiendo cada vez más claramente la sensación de pérdida irremediable, la misma sensación del niño que ve deslizarse hacia el cielo el globo que se le ha ido de la mano.

Está comprobado que bastan unas pocas vueltas para trastocar totalmente la organización del cubo. Que le sirva de advertencia al novato. Muchos principiantes tratan de ir afianzando sus progresos dejando lista primero una cara. Metidos ya en este atolladero dejan en cualquier sitio el cubo parcialmente resuelto, a la vista del primero que llegue. Tan pronto algún amigo lo coge, preguntando "¿Qué es esto?" se presenta en el aspirante a cubista el conocido síndrome de "¡no lo toques!", en alguna de sus formas más agudas. Empavorecido ante la idea de que la cara con tanto esfuerzo construida pueda quedar desmantelada grita "¡No lo toques!". Irónicamente, la victoria sólo puede alcanzarse con una actitud más flexible, que pasa precisamente por la destrucción parcial de lo ya logrado.

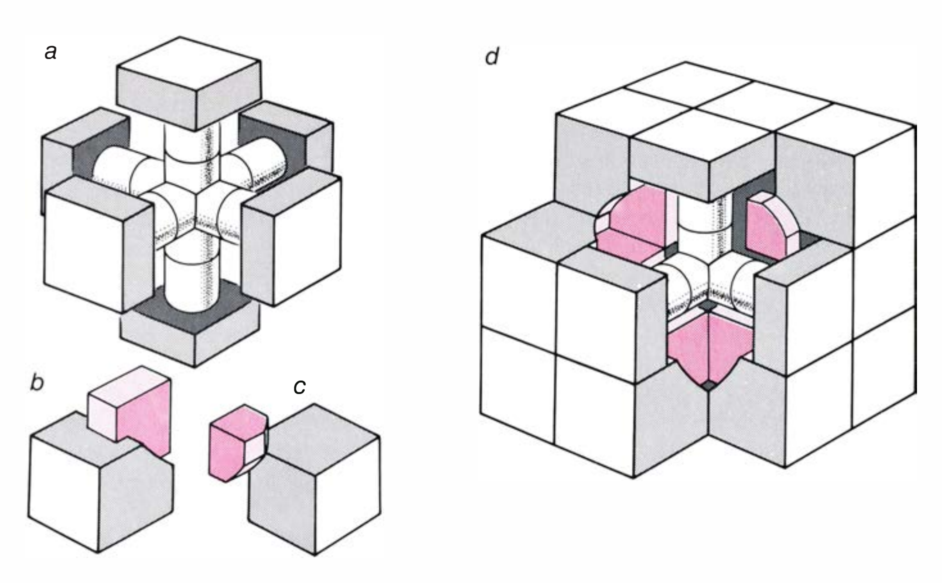
En el principiante, la destrucción del orden inicial comporta una ominosa sensación de irreversibilidad, un temor semejante al de caer al abismo. Cuando mi primer cubo (actualmente poseo cinco) quedó desordenado (por un invitado manitas) me sentí a un tiempo aliviado (puesto que era inevitable) y entristecido (temiendo que la organización inicial se hubiese perdido para siempre). El físico que llevo en mí empezó a pensar en la entropía. A partir del momento en que la posición inicial se había tornado inalcanzable, cada nuevo giro de una u otra cara parecía intrascendente. Mis ojos ingenuos no sabían distinguir un estado de desorden de otro, lo mismo que eran incapaces de distinguir una fuente de *spaghetti* de otra, un montón de hojarasca de otro, y así sucesivamente. Los detalles nada me decían y, por consiguiente, no eran tenidos en cuenta. Conforme iba realizando mi "caminata aleatoria" por el cubo, más y más vívidamente iba percibiendo la vastedad del espacio de posibles mezcolanzas de los cubículos.

Lo mismo que podemos contar todas las posibles barajaduras de un mazo de naipes, podemos calcular exactamente el número de posibles reordenaciones del cubo. Una estimación inicial se lograría como sigue. La primera cosa a observar –y bien elemental, por cierto– es que, al hacer girar una cara, cada vértice se transforma en otro, cada arista, en otra, y el cubículo central permanece invariable. Por consiguiente, los vértices tan sólo se mezclan entre sí, y lo mismo las aristas. Hay ocho piezas triédricas que ocupan los "cubículos" situados en los vértices, que ahora consideraremos como "nichos" recortados en el espacio, olvidando su contenido. Podemos llevar cada una de las piezas triédricas hasta cada uno de los ocho nichos situados en los vértices del cubo. Disponemos, pues, de ocho posibles ocupantes del nicho número 1, de siete candidatos para ocupar la celdilla número 2, de seis pares la número 3, y

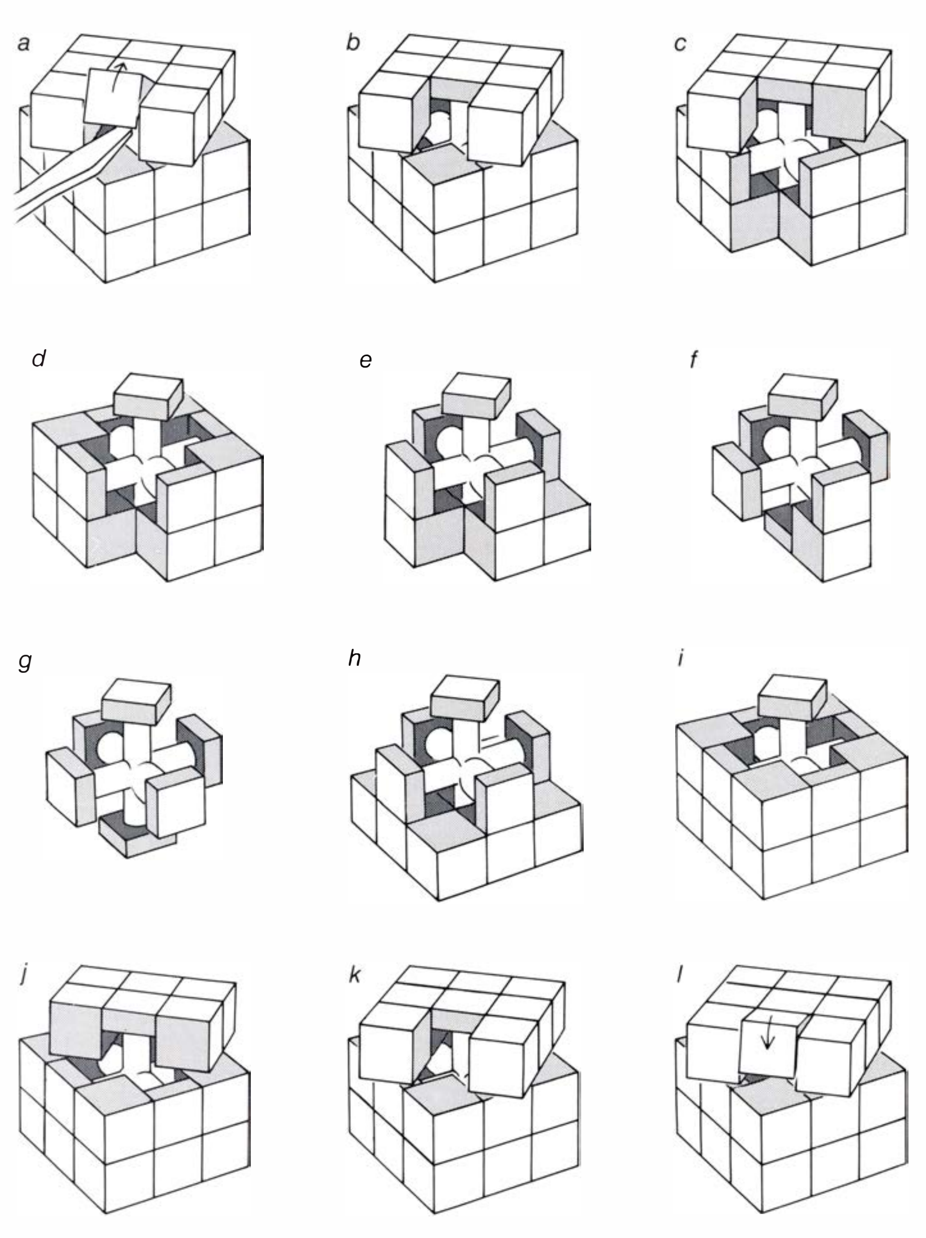
así sucesivamente. Por tanto, podemos alojar los vértices en sus cubículos de $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ (= 8!) formas diferentes. Empero, cada pieza triédrica puede estar dispuesta en una cualquiera de las tres orientaciones. Por consiguiente, deberíamos esperar que el número de permutaciones anterior viniera afectado del coeficiente 3^8. Análogamente, podríamos razonar para las piezas diédricas que forman las aristas: 12 objetos pueden permutarse entre sí de 12! maneras, y, dado que cada uno de éstos admite dos orientaciones, tendríamos que multiplicar este número por 2^{12}. Las facetas centrales nunca abandonan sus posiciones iniciales (salvo que el cubo sea volteado como un todo, lo que no cambia sus posiciones relativas), y no aportan nuevos factores. Calculando los productos obtenemos 519.024.039.293.878.272.000 posibles posiciones; vale decir, alrededor de 5×10^{20}.

Mas el cálculo anterior está basado en una hipótesis implícita, a saber, que cualquier pieza puede alcanzar cualquiera de los cubículos de su especie, y ello en cualquiera de las orientaciones posibles, independientemente de las posiciones y orientaciones de los demás. Como veremos, no sucede exactamente de este modo. Resulta que hay una ligadura entre las orientaciones de las piezas triédricas. Podemos orientar arbitrariamente a siete cualesquiera de ellas, pero la octava viene entonces obligada, lo que rebaja el número de posibilidades en un factor 3. Análogamente, hay también una restricción sobre las piezas de arista, diédricas: pueden orientarse, a capricho, 11 cualesquiera de las 12, mas la duodécima queda determinada por las anteriores, suprimiéndose así otro factor 2. Finalmente, hay otra limitación sobre las permutaciones de las piezas (con independencia de su orientación) que establece que podemos situarlas todas, menos dos, donde queramos, con la constricción de que las dos últimas quedan determinadas por las demás. Queda eliminado así un último factor 2, reduciendo en total la estimación inicial en el factor 12, con lo que las posibles presentaciones del cubo no pasan de unas meras 43.252.003.274.489.856.000, es decir, alrededor de 4×10^{19}. De todas formas, esta cifra supera "ligeramente" la declarada en la etiqueta del cubo "ideal": "Más de tres mil millones de combinaciones".

Hay otra forma de concebir este factor 12: que comenzando en la posición inicial "fundamental" sólo podemos alcanzar la doceava parte de las disposi-



*La cruceta (**a**), un cubito-arista (**b**), un cubito-vértice (**c**) y una vista del mecanismo interno (**d**)*

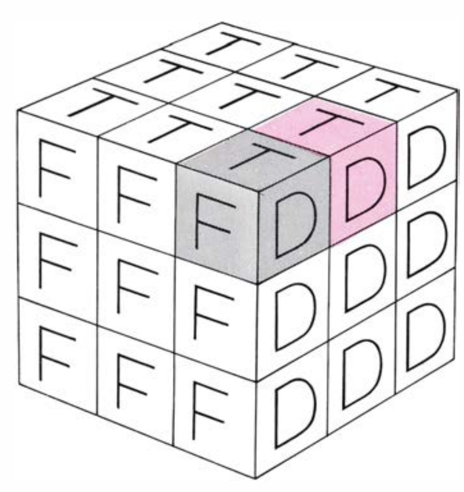


*Desmontaje (**a-f**) y ensamblado (**g-l**). ¡Nunca olvide restablecer la posición fundamental!*

ciones "evidentes". Si desarmamos el cubo y lo volvemos a montar con una sola de las piezas triédricas girada 120 grados respecto de su posición anterior, tendremos un "estado" inaccesible desde el fundamental, y este nuevo estado "genera" toda una nueva familia de 43.252.003.274.489.856.000 estados accesibles desde él. Hay en el cubo 12 familias mutuamente aisladas ("disjuntas") como las explicadas, familias que en teoría de grupos suelen llamarse "órbitas".

Y ya que hablamos de giros y torsiones imposibles, me gustaría mencionar un encantador paralelismo entre el cubo y la física de partículas elementales, que ha sido señalado por Solomon W. Golomb (a muchos lectores de Gardner este nombre les será sin duda familiar). Es imposible hallar una sucesión de giros que terminen por dejar uno de los cubitos de los vértices dislocado un tercio de vuelta, e invariantes a todas las demás piezas. En homenaje a la famosa partícula hipotética cuya carga ha de ser + 1/3, y de su antipartícula, de carga − 1/3, Golomb denomina *quark* al giro de un tercio de vuelta en sentido horario y, *antiquark,* al correspondiente giro de igual amplitud y sentido contrario. Al igual que sus homónimos del cubo, las partículas quark han mostrado ser esquivas hasta la exasperación, y muchos físicos teóricos sostienen actualmente el principio de confinamiento de los quarks, que afirma la imposibilidad de hallar un quark (o antiquark) libre, aislado. No deja de ser atractiva la correspondencia entre los quarks ligados al cubo y los ligados a partículas subatómicas.

En realidad, la relación entre ambos tipos va mucho más allá. Las partículas quark no pueden existir libres, pero sí agrupadas y ligadas entre sí. Un par quark-antiquark es un mesón, y un trío



Los cubitos **tdf (gris)** *y* **td (en color)**

de quarks cuya suma de cargas sea entera es un barión. (Ejemplo, el protón, cuya carga es + 1.) Ahora bien, en el Cubo Mágico, sorprendentemente, sí es posible darle a dos vértices giros de un tercio de vuelta, siempre que tales giros sean de *sentidos opuestos* (uno de sentido horario y otro de sentido antihorario). También podemos darle giros de un tercio de vuelta a tres vértices, con tal de que todos sean del *mismo* sentido. Por esta causa, Golomb llama *mesones* a los estados del cubo que presentan dos vértices vueltos en sentidos contrarios, y *bariones* a los que contienen tres vueltas en la misma dirección. En el universo de las partículas, tan sólo pueden existir combinaciones de quarks cuya carga neta sea expresable mediante enteros. En el universo del cubo, tan sólo son factibles combinaciones de quarks con giro total expresable en unidades completas. No es sino otra forma de decir que la orientación del octavo triedro viene siempre impuesta por las orientaciones de los otros siete. En el mundo del cubo, la razón subyacente al confinamiento de los quarks se encuentra en la teoría de grupos. No es imposible que exista una explicación semejante para el confinamiento de los quarks subatómicos, basada también en la teoría de grupos. La cuestión está por estudiar, pero, en cualquier caso, el paralelismo es tan grato como sugestivo.

Si tenemos un cubo prístino (cuyas caras se encuentran todas en posición fundamental), ¿de qué tipo son las series de movimientos capaces de crear mesones o bariones? Esta pregunta suscita una de las nociones más potentes de la cubología, a saber, sucesiones "enlatadas" de movimientos, capaces de ejecutar determinadas reordenaciones de unos cuantos cubos, dejando intacto todo lo demás (o como se dice en teoría de grupos, "invariante"). Estas series prefabricadas de movimientos han recibido muchas denominaciones. Las he oído llamar operadores, transformadas, palabras, herramientas, procesos, maniobras, rutinas, subrutinas y macros. Los tres primeros términos proceden de la teoría de grupos; los tres últimos son adaptaciones de la jerga de las ciencias de cómputo. Cada uno comporta un aroma característico, y a mi me gusta usar de todos ellos.

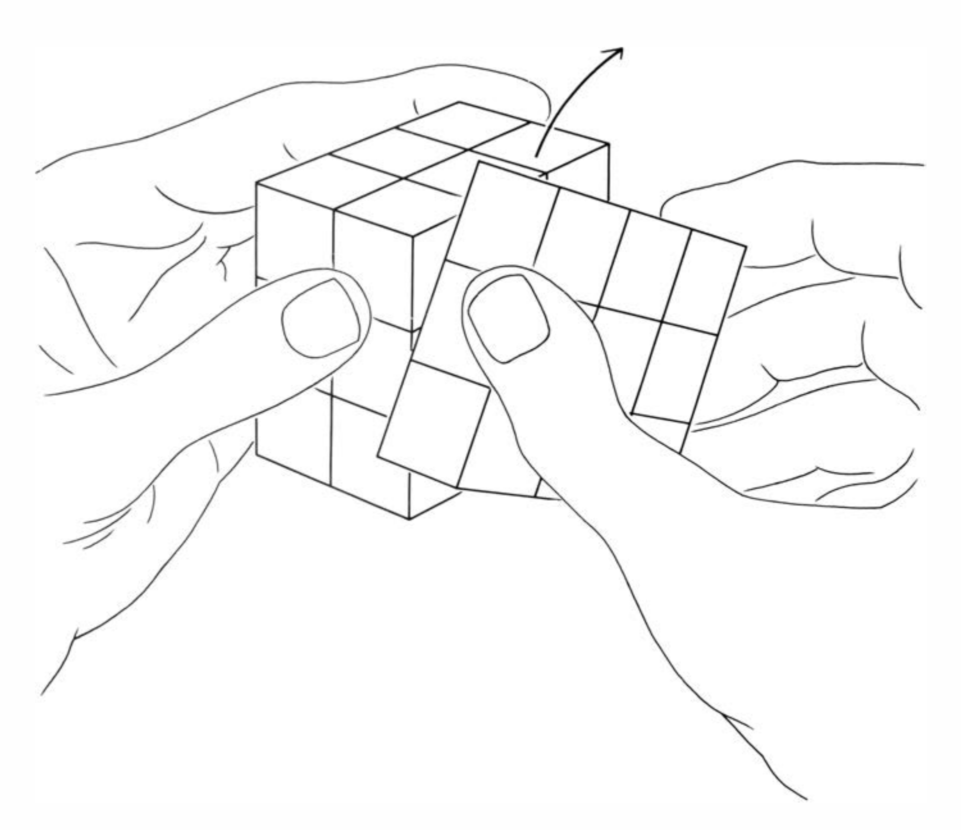
Hablar de procesos exige precisión, y para ello se requiere buena nomenclatura y técnica de notación. Presentaré ahora la ideada por Singmaster. Ante todo necesitamos una forma de referirnos a cada una de las caras del cubo. Cabe la posibilidad de darles a las caras los nombres de sus colores iniciales, incluso después de quedar entremezclados. Podría parecer absurdo decir que una cara es blanca, cuando tenemos manchas blancas salpicando todo el cubo. Recordemos, sin embargo, que la faceta blanca central nunca se mueve con respecto a las otras cinco centrales, y por ello sirve para definir la "cara blanca" del cubo. ¿Por qué no darles, pues, a las caras los nombres de los colores? El problema está en que las posiciones relativas de los colores suelen ser distintas en cubos diferentes. Incluso dos cubos fabricados por una misma marca pueden venir con posiciones fundamentales diferentes. Un convenio más general consiste, sencillamente, en nombrar las caras por su situación respecto a nosotros. Usaremos el siguiente sistema de iniciales: *I,* para la cara situada a nuestra izquierda; *D,* para la cara derecha. *F* será la cara situada frente a nosotros, y *A*, la situada en la parte de atrás. Finalmente, llamaremos *T* a la tapa, o cara superior, y *B* a la cara inferior, que sirve de base para el cubo. Con este convenio, podemos ahora designar cada uno de los cubitos componentes mediante grupos de letras minúsculas, que nos dicen las caras a que pertenecen. Así, por ejemplo, *td* (o *dt*) significa la pieza situada en la arista de las caras "tapa" y "derecha", mientras que *tdf* es el vértice situado en el triedro de las caras anteriores y la frontal. [*véase la ilustración de esta misma página*].

Para un cubista diestro, el giro más natural parece ser el mostrado en la ilustración. Vista desde el lado derecho, tal maniobra produce en la cara *D* un giro de un cuarto de vuelta en sentido horario. La imagen especular de esta maniobra, donde la mano izquierda haría girar la cara *I* en sentido antihorario (vista desde el semiespacio izquierdo) es I^{-1}, o más brevemente, *I'*. El giro de 90 grados en sentido horario de la cara *I* se llamará, por analogía con la definición del giro *D,* al igual que la cara que gira, es decir, *I.* En general, el giro necesario para rotar una cara 90 grados en sentido horario (respecto de un observador situado frente a ella) se denota por la letra mayúscula que da nombre a la cara, y su inverso –giro de igual amplitud y sentido antihorario– se designa colocando tras la inicial correspondiente un apóstrofe (') o el exponente $^{-1}$. En lo sucesivo, los giros de amplitud un cuarto de vuelta se llamarán "cuartos de giro", o, brevemente, *q*-giros.

Provistos de esta nomenclatura, ya es posible describir cualquier sucesión de movimientos, por compleja que resul-

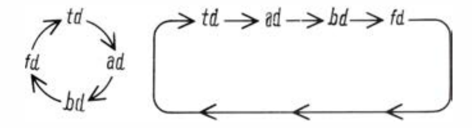
te. Un ejemplo trivial: cuatro *q*-giros *D* consecutivos, que expresamos D^4. Con vocabulario de teoría de grupos, esta operación es el neutro o identidad del grupo; nosotros lo denotaremos *0*, porque su efecto es nulo. La igualdad que así lo expresa es $D^4 = 0$.

Hagamos ahora girar dos caras, por ejemplo, primero *D* y luego *T*, sendos cuartos de vuelta. Denotaremos tal acción por *DT* y no *TD*. Observemos ante todo que los efectos de *TD* y *DT* son completamente diferentes. Para comprobarlo, ejecutemos primero *DT* sobre un cubo prístino, fijémenos en el resultado, y desandemos el camino. Ensayemos ahora *TD*, y apreciaremos la diferencia de efectos. Está claro que el movimiento inverso de *DT* es *T'D'*, y no *D'T'*. (Incidentalmente, la táctica de ir ensayando en el cubo prístino secuencias de movimientos es de la máxima utilidad. Muy pronto descubrí la ventaja de adquirir un segundo cubo, pudiendo así trabajar en la resolución de uno mientras tanteaba con el otro, sin apartarme mucho de la posición inicial.)



El movimiento más natural para el cubista diestro es **D**

¿Qué efectos produce una "palabra" determinada? Es decir, ¿qué facetas se mueven, y adónde? No podemos responder sin dar un método para expresar los desplazamientos de los cubitos individuales. El efecto de *D* sobre las aristas es llevar la pieza *td* hasta la cara de atrás, donde ocupará el cubículo *ad*. Al mismo tiempo, el cubito situado en *ad* pasará a la cara inferior, aterrizando en *bd;* el situado en *bd* avanza en el carrusel, saltando a *fd,* y el aquí situado asciende hasta *td.* Simbólicamente podemos expresar todo el proceso de una de estas formas:



Llamaremos ciclos de orden 4, o simplemente 4-ciclos, a estos "carruseles", que denotaremos más concisamente así: (*td, ad, bd, fd*). Desde luego, poco importa por donde lo empecemos; podríamos igualmente escribir (*bd, fd, td, ad*). Sí es importante, en cambio, el orden de las letras en los nombres de los cubitos. Si queremos, podemos invertir el orden de las letras en todos ellos; podemos también no permutarlas en ninguno. Pero no podemos hacerlo en unas sí y en otras no. Para comprenderlo basta recordar que las letras designan facetas. Escribiendo, por ejemplo (*td, da, bd, df*), estaríamos expresando un 4-ciclo donde intervienen los mismos cuatro cubículos, pero ahora cada vez que un cubito pasa de un cubículo (nicho) al siguiente tiene que "virar" (con la acepción que usan los químicos), es decir, cambiar la orientación de los colores de sus caras. Es evidente la imposibilidad de lograr semejante ciclo mediante un único *q*-giro, pero sí pudiera ser resultado de una *sucesión* de *q*-giros de caras distintas, es decir, de un *operador.* Fijémonos si no en el 8-ciclo siguiente: (*td, tf, ti, ta, dt, ft, it, at*). Aunque de longitud ocho, sólo intervienen en él cuatro cubículos. Cada cubito, tras darle una vuelta a la tapa, retorna virado; tras una segunda vueltecita queda como estaba al empezar. Cada faceta describe, por así decirlo, una "finta de Möbius". Al "4-ciclo virado" anterior podríamos denotarlo $(td, tf, ti, ta)_{+}$, donde el signo "más" denota los virajes. Sería igualmente posible denotarlo $(dt, ft, it, at)_{+}$, o de otras muchas maneras. La notación cíclica no sólo nos dice hasta dónde se desplaza el cubito, sino también su orientación con respecto de los otros cubitos de su ciclo.

Para terminar la descripción del efecto de *D,* tendremos que transcribir también el 4-ciclo de los vértices. Al igual que con las aristas, tenemos libertad para empezar por el vértice que más convenga, siendo también preciso seguir la pista al desplazamiento de las facetas, dando correctamente sus orientaciones. Empero, el efecto de *D* sobre los vértices es bastante trivial: (*tdf, adt, bda, fdb*), que por supuesto podríamos escribir de muchas otras formas. En resumen, podemos escribir *D* = (*td, ad, bd, fd*) (*tdf, adt, bda, fdb*), igualdad que nos dice que *D* está formado por dos ciclos disjuntos. (De haberlo querido, hubiéramos podido introducir un término que representase la rotación de 90 grados de la faceta central de la cara *D,* pero al ser indetectable su rotación, no lo añadiremos.)

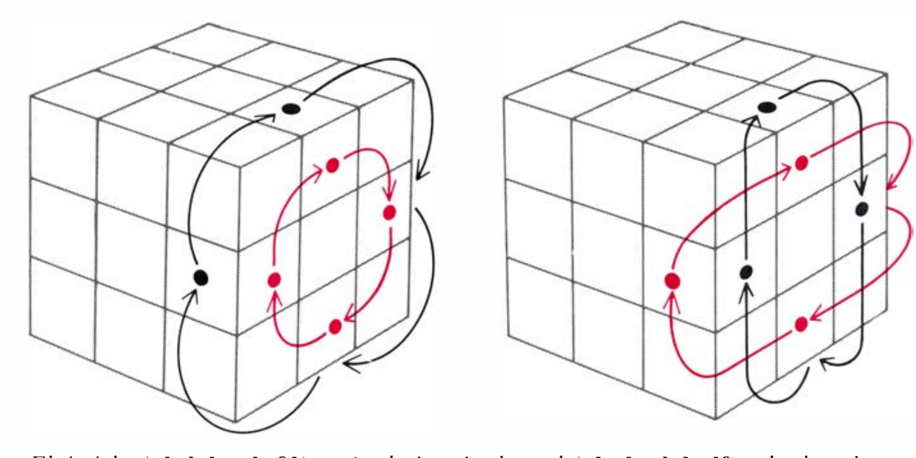
¿Qué transcripción dar de una sucesión de movimientos, tal como *DT*? Bien, tomemos un cubo prístino y ejecutemos sobre él *DT*. Empecemos entonces con cualquier cubito que se haya movido; describamos su trayectoria. Por ejemplo, *td* se ha mudado a *ad*. Por consiguiente, el antiguo inquilino de *ad* ha sido desplazado. ¿Adónde ha ido? Busquemos la nueva dirección de este cubito (ahora es *bd*) y continuemos siguiéndole la pista a otros cubitos en sus periplos en torno al cubo, hasta descubrir cuál ha venido a ocupar el alojamiento inicial de *td.* Se descubrirá el siguiente 7-ciclo: (*td, ad, bd, fd, tf, ti, ta*). [*Véase la ilustración de la página 129*].

¿Y qué hay de los vértices? Supongamos que le seguimos la pista al cubito que partió de *tdf.* ¿Adónde lo trasladó *TD*? La respuesta es: a ningún sitio; este vértice dio una vuelta en torno al cubo sufriendo una torsión conforme

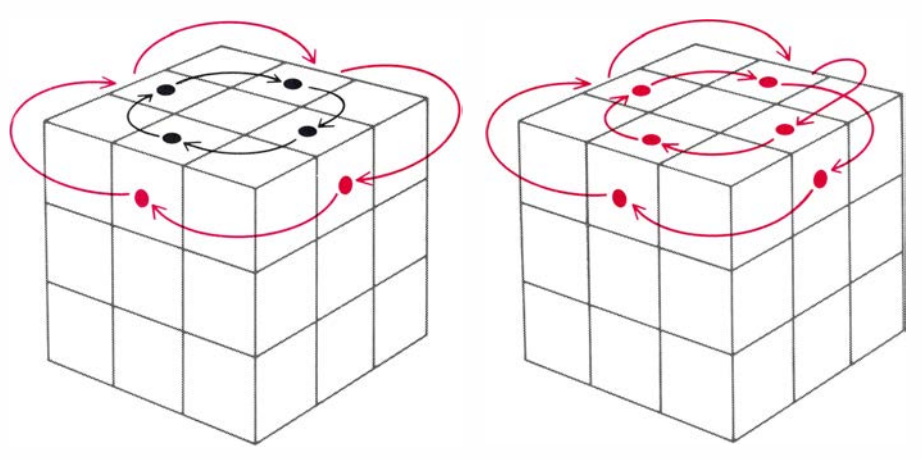
describía el circuito, ya que se ha transformado en *dft.* Podemos denotar $(tdf)_+$ a esta torsión de sentido horario –un quark positivo–. Este "uniciclo de torsión" puede servirnos para abreviar el 3-ciclo siguiente: (*tdf, dft, ftd*). Como vemos, en este 3-ciclo se van permutando circularmente las letras *t, d, f* del nombre del cubito. Si el ciclo hubiese sido un antiquark lo habríamos denotado $(tdf)_-$, y las letras irían permutándose circularmente en sentido contrario.

¿Qué podemos decir de los otros siete vértices? Dos de ellos –*bai* y *bif*– permanecen inmóviles, mientras que los cinco restantes *casi* forman un 5-ciclo: (*tad, abd, bfd, itf, ati*). Es una lástima que el ciclo no se cierre del todo. En efecto, aunque *ati* se transfiere al cubículo inicial *tad,* sufre torsión por el camino, finalizando realmente en *dta,* que dista de *tad* una torsión de sentido antihorario. El ciclo verdadero tiene longitud 15, pero al estar tan cercano al 5-ciclo anterior nos decidimos por subindicar éste con un signo menos que denote la torsión de sentido antihorario. El ciclo de torsión de orden 5 es, pues, $(tad, abd, bfd, itf, ati)_-$, y el efecto completo de *DT,* expresado mediante ciclos, es (*td, ad, bd, fd, tf, ti, ta*) $(tdf)_+$ $(tad, abd, bfd, itf, ati)_-$.

Una vez descompuesto *DT* en producto de ciclos podemos ejecutar rotaciones mentalmente, por puro cálculo. Por ejemplo, ¿cuál sería el efecto de $(DT)^5$? El cubito-arista *td* sería llevado cinco pasos hacia adelante en su ciclo, quedando finalmente en *ti.* (Podemos también concebir que *td* retrocede dos pasos.) Análogamente, *ti* se trasladaría a *fd,* y lo mismo los demás. El 7-ciclo anterior quedaría reemplazado por el nuevo: (*td, ti, fd, ad, ta, tf, bd*). Examinemos ahora el 5-ciclo de torsión. El cubito del triedro *tad* sería promocionado cinco puestos en su ciclo, regresando así al punto de partida, si bien con torsión negativa, es decir, convertido en *dta.* Análogamente, todos los cubitos triedros mencionados en el 5-ciclo retornarían a sus alojamientos iniciales, si bien retorcidos negativamente; por tanto, al elevar a la quinta potencia un 5-ciclo de torsión negativa éste se desintegra en cinco antiquarks. Mas, de ser esto cierto, ¿de qué forma se respeta el principio de cuantificación entera de la torsión total? ¿No es cierto, acaso, que tenemos un quark $-(tdf)_+-$ y cinco antiquarks, cuya suma es cuatro antiquarks, con torsión total $-1^1/_3$? Bueno, la verdad es que con las prisas me he dejado algo en el tintero. ¿Sabría el lector detectar qué ha sido? Para adquirir soltura con la notación cíclica, el lector debería esforzarse en calcular la representación de las diversas potencias de *DT* y *TD,* así como las de sus inversas.



*El 4-ciclo (***td, bd, ad, fd***) está a la izquierda y el (***td, da, bd, df***) a la derecha*



*A la izquierda, el 4-ciclo (***td, tf, ti, ta***).*
*A la derecha, el 4-ciclo (***td, tf, ti, ta***)$_+$ con virado*

Toda sucesión de movimientos es representable mediante ciclos disjuntos de longitudes adecuadas. Dos ciclos son disjuntos cuando carecen de elementos comunes. De todas formas, si se admite que los ciclos puedan compartir elementos, todo ciclo es ulteriormente descomponible en producto de 2-ciclos (cuyo nombre técnico es transposiciones, pero que llamaremos también trueques). Imaginemos, por ejemplo, tres animales: un caimán, un lince y un dromedario, que inicialmente viven en otros tantos *habitat* ecológicamente idóneos, que aquí denotaremos *A, B* y *C.*

El efecto del ciclo ternario (*A, B, C*) es alojarlos en el orden dromedario, caimán, lince. El mismo resultado puede conseguirse ejecutando primero la transposición (*A, B*), que lleva a *B* lo que haya en *A,* y recíprocamente y, seguidamente, la transposición (*A, C*). Evidentemente, también se consigue igual resultado con las transposiciones sucesivas (*A, C*) (*B, C*) o las (*B, C*) (*A, B*). Por otra parte, ninguna secuencia de tres transposiciones produce iguales efectos que (*A, B, C*). El lector debería hacer los tanteos y razonamientos necesarios, y convencerse por sí mismo. (Observemos que los hábitat hacen el papel de cubículos, y los animales de cubitos.)

Un teorema elemental de la teoría de zoológicos (campo donde no haremos incursión alguna) establece que, dada una permutación cualquiera de los animales entre los distintos hábitat, siempre es posible reducirla a un producto de transposiciones consecutivas. Aunque tal descomposición no es única, el número de trueques siempre tiene la misma paridad; es decir, si por un procedimiento se ha descompuesto la permutación en un número *par* de transposiciones, por ningún otro método podrá ser descompuesta en número *impar.* Además de ser invariante, la paridad de cada permutación es suma de las paridades de cualesquiera permutaciones en que se descomponga (usando las reglas de adición de números pares e im-

pares: par más impar da impar, etcétera).

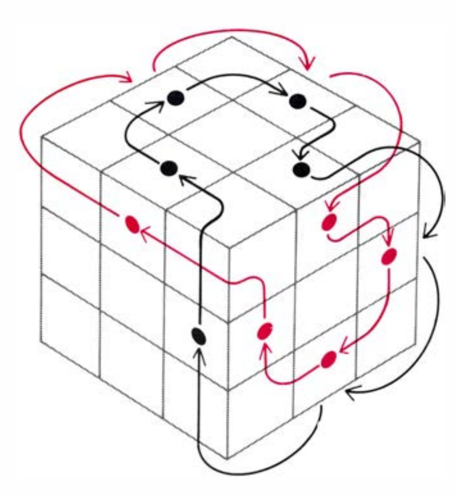
Este teorema tiene repercusión en el Cubo Mágico. En particular, se puede comprobar que todo q-giro está formado por dos 4-ciclos disjuntos (uno correspondiente a vértices, otro, a las aristas). ¿Cuál es la paridad de los q-giros? Como el lector puede comprobar, es impar. Por tanto, tras cada q-giro los vértices y aristas sufren sendas permutaciones impares; tras dos q-giros, permutaciones pares; después de tres q-giros habrán experimentado permutaciones impares, y así sucesivamente. Vértices y aristas se conservan en fase, en el sentido de que las paridades de sus permutaciones siempre son idénticas. Entonces, es evidente que la permutación nula o "identidad" es de clase par (cero transposiciones). Por consiguiente, si tenemos una permutación nula en los vértices, la clase de permutación de aristas también ha de ser par. Recíprocamente, tener en las aristas la permutación nula implica tener en los vértices una permutación de clase par. Imaginemos un estado idéntico al inicial salvo en dos aristas, que se encuentran intercambiadas (es decir, ha habido un trueque de posiciones entre ellas). Tal organización del cubo es par en los vértices e impar en las aristas: resulta, pues, imposible. Lo más parecido que podemos lograr es *dos pares* de aristas trastocadas. El mismo razonamiento vale para los vértices. En resumen: son imposibles los trueques aislados; deben siempre presentarse por pares. (Este es el origen de uno de los divisores "dos" que introdujimos al calcular el número de posibles estados del cubo.) Hay procesos para intercambiar dos pares de aristas, dos pares de vértices, e incluso, un par de vértices juntamente con un par de aristas. (Este último proceso requiere obligatoriamente un número impar de q-giros.)

Para dejar totalmente lista la cuestión de las ligaduras, reflexionemos acerca del origen de las restricciones sobre torsión de vértices y pivotado de las aristas. He aquí una perspicaz explicación debida a los matemáticos John Horton Conway, Elwyn R. Berlekamp y Richard K. Guy, construida sobre ideas de Anne Scott. La idea a tener presente es que deseamos probar que el número de cubitos diédricos que han sufrido viraje siempre es par, y que la torsión total de los vértices es un número entero. Para poder determinar si se ha volteado o retorcido alguna pieza es imprescindible un sistema de referencia. A tal fin definimos las nociones de "faceta principal de un cubículo" y de "color principal de un cubito". (Recordemos que los cubículos son nichos y que los cubitos son objetos sólidos.) La faceta principal de un cubículo será la situada en la tapa superior o en la cara base, si en ellas tiene alguna; en los demás casos será la faceta situada en la pared derecha o en la izquierda [*véase la figura superior de la ilustración de la página 131*]. Hay nueve facetas principales en la cara *T*, y otras nueve en *B*, más cuatro situadas en el ecuador. Podemos olvidarnos de los centros, pues nunca sufren torsiones o virados. El color principal de un cubito es, por definición, el color que debería tener la faceta principal del cubito cuando éste "volviera a casa", y quedase alojado en el cubículo adecuado de la posición inicial.

Supongamos ahora que el cubo esté hecho un revoltillo. Todos los cubitos que tengan su color principal en la faceta principal del cubículo que están ocupando serán tenidos por cuerdos, pues "están en sus cabales". Los demás están "virados" o "flipados" (si son cubitos-arista) o "retorcidos" (cubitos-vértice). Está claro que un cubito puede haber sufrido torsión de dos formas: en sentido horario (+ 1/3 de vuelta) o en sentido antihorario (− 1/3 de vuelta). El "revirado" total de un estado del cubo se define como el número de cubitos virados que contiene; la "torsión total" es la suma de las torsiones de los ocho cubitos triédricos. Por convenio, el revirado y la torsión del estado inicial son ambos nulos.

Fijémonos ahora en los 12 q-giros que por composición generan todos los movimientos. Al realizar *T* o *B* (o sus inversos) se preservan tanto el revirado como la torsión, pues ningún cubito entra o sale de las caras tapa o base. Al ejecutar *F* o *A* (o sus inversos) la torsión permanece constante, pues cambia simultáneamente la torsión de cuatro vértices: dos aumentan + 1/3, y dos pierden − 1/3 cada uno. También deja invariable el revirado [*véase la figura central de la ilustración de la página 131*]. Al realizar *I* o *D* permanecerá constante la torsión (cuatro vértices sufren torsiones que se cancelan por pares) mientras el viraje aumenta en 4, pues en ambos casos hay cuatro cubitos que pivotan [*véase la figura inferior de la página 131*]. La conclusión fue ya enunciada sin demostración: la orientación de los ocho cubitos triédricos ha de ser tal que la suma de torsiones sea entera, mientras que los 12 cubitos diédricos siempre han de tener número par de virajes.

Tras este análisis de las restricciones del cubo, el lector debería haber quedado convencido de que, por mucho que se tuerza y retuerza el Cubo Mágico, desde la posición fundamental no podrá alcanzarse más de la doceava parte del "universo" imaginable. Cuestión completamente distinta, empero, es poder demostrar que todos los estados de este doceavo de universo son realmente accesibles desde la posición inicial (o lo que es equivalente: que el estado fundamental es accesible desde todos los del doceavo de universo). Para ello es necesario mostrar cómo pueden lograrse todas las permutaciones de clase par de los cubitos, y cómo pueden alcanzarse todas las orientaciones que no contravengan las limitaciones ya conocidas. Lo que se reduce a demostrar la existencia de operadores capaces de ejecutar siete tipos de procesos: (1) trueque simultáneo de dos pares de aristas cualesquiera, (2) trueque simultáneo de dos pares de vértices cualesquiera, (3) viraje simultáneo de dos pares de aristas, (4) un mesón arbitrario, (5) cualquiera de los 3-ciclos de aristas, (6) cualquiera de los 3-ciclos de vértices y (7) un barión arbitrario.



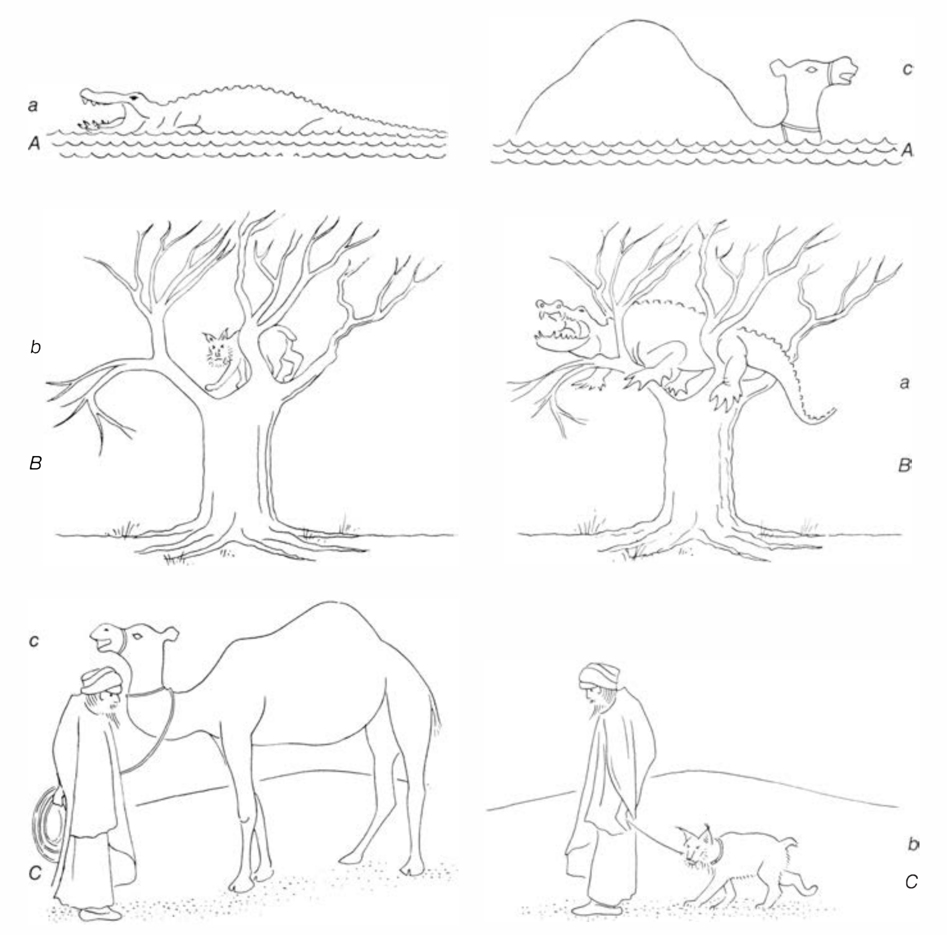
*El 7-ciclo (***td, ad, bd, fd, tf, ti, ta***)*

Evidentemente, cada uno de estos operadores debe lograr su fin sin provocar efectos secundarios en ninguna otra parte del cubo. Dotados de tan poderoso instrumental seríamos capaces de cubrir sin dificultad el doceavo de universo. En el caso de las permutaciones de fieras demostré ya que un 3-ciclo se descompone en 2-ciclos con un elemento común; lo que implica que las operaciones de tipos 5 y 6 pueden en realidad generarse por las cuatro primeras. Análogamente, los bariones pueden sintetizarse a partir de dos mesones no disjuntos. Así pues, todo lo que en realidad necesitamos son las cuatro primeras clases de operadores.

Para demostrar que están a nuestro

alcance los operadores de estas cuatro clases, nos apoyaremos en otra de las más cruciales y encantadoras nociones de la cubología: la conjugación de elementos. Resulta que tan sólo es necesario disponer de un ejemplar de cada tipo; a partir de él podemos construir todos los demás operadores de su misma clase. ¿Cómo es esto posible? La idea es muy sencilla.

Supongamos descubierto ya un operador de la clase (1), que transpusiera, por ejemplo, *tf* con *ta* y *ti* con *td,* dejando intacto el resto del cubo [*véanse las flechas de color de la ilustración de la página 132*]. Llamemos *J* a tal operador. Supongamos ahora que se quieran intercambiar dos pares de cubos diédricos enteramente diferentes, por ejemplo, *df* con *bf* y *da* con *bd* [*véanse las flechas negras de la misma ilustración*]. Dejando volar nuestros deseos, podríamos soñar: "Si tuviéramos la suerte de que los cubitos éstos se encontraran en los cuatro puntos de 'transposición mágica' de la tapa superior..." Bueno, ¿y qué nos impide llevarlos allí? Sería bien sencillo lograr que cuatro cubitos dados llegasen hasta cuatro cubículos prefijados. La pega es evidente: "Sí, pero a costa de embrollar terriblemente el resto del cubo". Empero, para esta objeción hay una salida muy airosa. Llamemos *A* a la maniobra –no importa cuán destructiva– que lleva nuestros cubitos hasta los puntos de transposición mágica. Supongamos que hemos ido anotando la secuencia de movimientos que componen *A.* Inmediatamente después de ejecutar *A* llevamos a cabo la operación de trueque doble, *J.* Ahora viene la parte más lúcida. Leyendo en orden inverso la lista de instrucciones para efectuar *A,* y ejecutando en sentido contrario cada uno de los giros allí anotados, estaremos materializando exactamente la maniobra inversa de *A.* Con ello no sólo devolveremos los cuatro cubitos a sus antiguos cubículos, sino que también desembrollaremos la tremenda confusión creada por *A.* ¿Quedará entonces el cubo rigurosamente intacto? No del todo. Al estar *J* emparedado entre *A* y *A',* los cuatro cubitos retornan trastocados, es decir, cada uno de ellos acaba alojándose en el antiguo domicilio de su "partenaire" de cambalache. En todo lo demás, el cubo recupera su organización anterior. Hemos conseguido precisamente el doble intercambio que nos proponíamos lograr.



*Un caimán, un lince y un dromedario (***a, b, c***) sufren una permutación de sus habitat ecológicos (***A, B, C***)*

Analizándolo con calma, percibimos la impecable concepción del razonamiento anterior. La maniobra inversa *A'* no puede "saber" que hemos intercambiado dos pares de aristas. Por lo que a ella se refiere, su trabajo consiste en devolver todo al estado en que se encontraba antes de ser ejecutada *A.* Con símbolos, hemos ejecutado la sucesión de movimientos *AJA';* el movimiento que produce se llama "conjugado" de *J.*

Es precisamente este tipo de ilustraciones maravillosamente concretas de nociones abstractas de la teoría de grupos la causa de que el Cubo Mágico sea uno de los más apasionantes instrumentos jamás ideados para facilitar la enseñanza de nociones matemáticas. Por lo común, en los cursos sobre teoría de grupos los ejemplos de conjugación que se presentan o son demasiado triviales, o demasiado abstrusos para estimular la imaginación o aclarar el concepto. El Cubo Mágico nos proporciona vívida ilustración del concepto de elementos conjugados y de otras muchas nociones importantes de la teoría de grupos.

Supongamos que se desee conseguir un par quark-antiquark en vértices diagonalmente opuestos, pero que sólo sepamos hacerlo cuando los vértices sean contiguos. ¿Cómo lograrlo? He aquí una pista: hay dos soluciones elegantes; la más breve y armoniosa requiere un proceso de conjugación. Incidentalmente, toda maniobra capaz de crear un quark en un vértice (con otros efectos secundarios, desde luego) será llamada "quarkillador".

Lo que hemos demostrado para aristas vale igualmente para vértices: la posibilidad de transponer dos vértices *específicos* permite transponer dos vértices *cualesquiera.* La conjugación permite construir toda una clase de operadores a partir de uno cualquiera de los miembros de esa clase. Obviamente, queda por resolver la cuestión de cómo hallar un operador "representante" en cada una de las cuatro clases. Por ejemplo, ¿cómo se da con un operador que sintetice un mesón sobre dos vértices adyacentes (una combinación de quarkillador y antiquarkillador)? ¿Cómo atinar con un operador capaz de intercambiar dos pares de aristas situadas sobre la tapa superior? No daré aquí las soluciones, sino que, siguiendo el criterio de Singmaster, indicaré la conveniencia de explorar cuasi-sistemáticamente ciertos "sub-universos" pequeños, seleccionados de entre la totalidad de estados del cubo. Más precisamente, Singmaster sugiere el examen de *sub-*

grupos, lo que significa limitar deliberadamente a ciertos tipos especiales el conjunto de movimientos admisibles. He aquí cinco ejemplos de subgrupos interesantes creados al imponer distintos tipos de restricciones:

1. Grupo de las secciones centrales. En este subgrupo cada giro de una cara debe ir acompañado del movimiento paralelo sobre la cara opuesta. Por tanto, D irá acompañado de I', T irá conjuntamente con B', y F, inmediatamente seguido de A'. El nombre del grupo procede de que ejecutar cualquiera de estos movimientos dobles es equivalente a girar alguna de las secciones centrales del cubo. Con la notación de Singmaster, el movimiento compuesto DI' se denota D_s; $D'I$ es D'_s, y así sucesivamente. Con esta restricción es imposible entremezclar arbitrariamente las caras. En todas ellas se reconocerá una pauta característica, pues sus vértices son siempre del mismo color. Como caso particular tenemos el motivo llamado "Lunares", donde cada cara es toda de un color, a excepción del centro [*véase la ilustración de la página 123*]. ¿Tiene el lector idea de cómo lograr el cubo de lunares partiendo de la posición fundamental? ¿De cuántas maneras distintas pueden estar repartidos los lunares? ¿En qué podemos decir que se parecen los mesones y la configuración de lunares? (Nota: El lector podrá encontrar las soluciones, y mucha más información, en el libro de Singmaster. Por favor, no me envíen soluciones. Remítanle a Singmaster, en cambio, las ideas nuevas.)

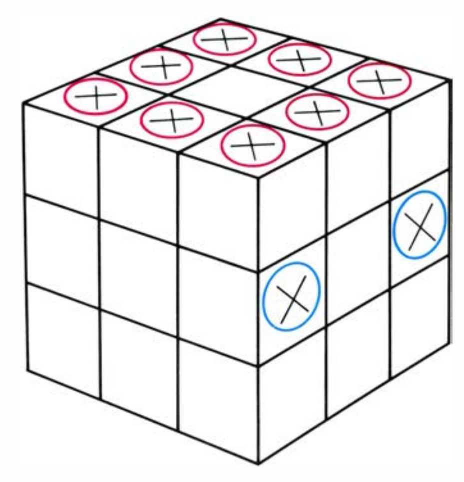
2. El grupo de secciones centrales al cuadrado. Ahora restringimos todavía más el grupo de secciones centrales, permitiendo tan sólo ejecutar cuadrados de los movimientos de aquél, como por ejemplo, D^2_s (que es igual a D^2I^2) o F^2_s (que es idéntico a $F^2 A^2$).

3. El grupo seccional antideslizante. Ahora, en lugar de hacer que las caras opuestas giren paralelamente, se las gira antiparalelamente, con lo que D va acompañado de I, F de A y T va seguido de B. Los movimientos antideslizantes llevan subíndice *a*. Así, D_a significa DI. (Evidentemente, el grupo antideslizante al cuadrado es idéntico al grupo seccional al cuadrado.)

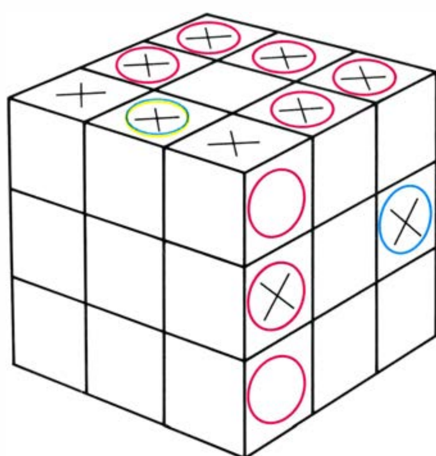
4. El grupo bifacial. Tan sólo es lícito ahora girar dos caras, siempre las mismas, por ejemplo, F y D.

5. El grupo bicuadrático. Al igual que en el grupo bifacial, tan sólo se pueden hacer girar dos caras, pero ahora los giros han de ser de 180 grados. Es, evidentemente, un grupo muy sencillo.

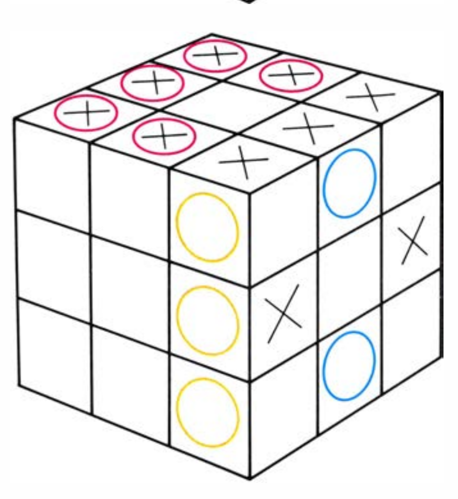
El cubo está en posición fundamental. Las facetas principales de los cubículos están señaladas con *X* negras, los colores principales, por *O* de color. (Podemos prescindir de los centros, pues permanecen fijos.) Debemos considerar las *X* flotando en el espacio y, las *O*, ligadas al cubo. Así, al moverse las caras, las *X* permanecerán donde están, mas las *O* pasarán a ocupar otros lugares.



Se ha ejecutado el *q*-giro *F*. Los dos círculos vacíos muestran que esas dos facetas han "perdido el juicio". Para devolverles la cordura tendríamos que girar un tercio de vuelta en sentido horario y otro tercio, en sentido contrario. En la cara izquierda, no visible, ha sucedido otro tanto.



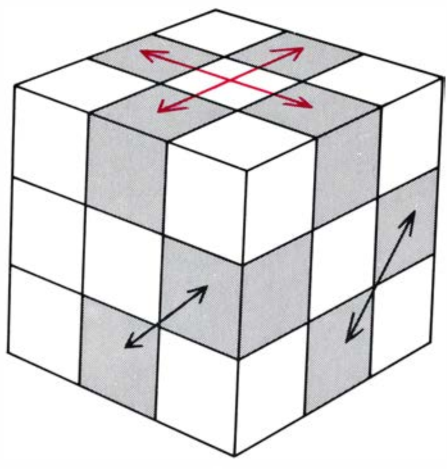
Partiendo de la posición fundamental se ha ejecutado el *q*-giro *D*. Los círculos vacíos se presentan por pares. Los amarillos de los vértices superior y básico sufren torsiones mutuamente destructivas. Las aristas azules forman un par; el borde amarillo está emparejado con otro, invisible, situado en la cara de atrás, que ha sufrido viraje.



Demostración de que el revirado es par y, la torsión, entera

Centrando la atención tan sólo en los grupos bifaciales y bicuadráticos, se podrán descubrir procesos capaces de efectuar dobles trueques, de aristas en unos casos, de vértices en otros. Es muy notable que –con auxilio de la conjugación– estos dos procesos permitan, al menos desde el punto de vista teórico, resolver siempre el problema completo de reordenación.

¿Por qué podemos salir adelante sin un generador de mesones y sin un generador específico para virar pares de aristas? Veamos cómo podríamos construir este último proceso a partir de los dos tipos de herramientas que suponemos ya obtenidas. Para pivotar dos cubitos diédricos sin provocar efectos secundarios, realizaremos consecutivamente dos trueques de dobles pares de aristas, ¡que harán intervenir las dos veces a los mismos pares! Por ejemplo, podríamos transponer *tf* con *ta* y *bf* con *ba*, y volver seguidamente a transponer estos mismos pares. Parece que este proceso será "absolutamente nulo", pero no es forzoso que sea así. Después de todo, como ya hemos hecho antes,



Transposición de dos pares de aristas cualesquiera

podemos emparedar la segunda transformación entre un proceso X y su inverso X', eligiendo cuidadosamente el proceso X para que... (¡Vaya por Dios! ¡He perdido completamente el hilo del razonamiento! De todas formas, estoy seguro que los lectores podrán terminarlo sin dificultad. Recuerdo que no era demasiado difícil, y que la idea me había parecido elegante. Creo que los lectores opinarán igual.)

El mismo tipo de consideraciones mostraría cómo se puede construir un generador de mesones partiendo de meras transposiciones de vértices convenientemente aderezadas de operaciones de conjugación. Disponiendo de mesones podemos sintetizar bariones. Y con bariones y mesones, viradores de dobles parejas de aristas, transmutadores de dobles parejas de aristas y dobles parejas de vértices, se dispone de instrumental suficiente para restaurar en el cubo su perfecta organización inicial partiendo de cualquiera de sus estados desordenados, siempre y cuando pertenezcan a la misma órbita que el estado fundamental. Inútil es decirlo, la demostración de existencia que acabamos de presentar es sumamente teórica; cualquier sistema operativo de rutinas capaces de resolver el cubo tendrá que estar organizado de muy otra manera. El tipo de solución que hemos presentado tiene la ventaja de admitir descripción sucinta, pero desde el punto de vista práctico es tremendamente ineficaz. Para tener éxito, el cubista tiene que poner a punto un sistema de rutinas tan extenso como flexible, fácil de memorizar y altamente redundante. Es una gran ventaja poder realizar de diversas formas las transformaciones necesarias: se puede entonces optar por la herramienta que mejor parezca adaptarse a la situación que se tenga entre manos, en lugar de emplear, pongamos por caso, algún instrumento teórico capaz, sí, de generar un barión tras varios centenares de q-giros.

El cubista típico llega a poner a punto un sistema de transformaciones en parte por intuición, en parte por suerte, a veces ayudándose de esquemas y diagramas y, en ocasiones, con ayuda de principios abstractos de teoría de grupos. Uno de los principios que casi todo el mundo enuncia muy pronto es el de "ir apartando estorbos", que no es sino el principio de conjugación presentado de forma intuitiva. En su aplicación práctica, el principio suele ir acompañado de un monólogo (al que he añadido efectos sonoros a propósito): "Vamos a ver, sacaré *esto* de enmedio (flip, flip) para poder mover *esto otro* (flap, flap), y ahora ya puedo devolver el de antes donde estaba (desflip, desflip). Bueno, ya tengo a *esto* donde yo quería". No hay duda: la letra y la música son de "la canción del conjugado".

Tener conciencia de las causas que hacen que todo vaya saliendo bien tiene una pega: puede resultar demasiado fatigoso. Tengo la impresión de que muchos cubomaestros no analizan con tanto detalle la causa de que sus herramientas vayan produciendo los efectos deseados, por lo menos, no mientras están a medio camino de restaurar un cubo en revoltillo. Los cuberos expertos son más bien como virtuosos del piano, que han aprendido de memoria pasajes muy difíciles. Tal como dijo un cubomaestro del Instituto de Tecnología de Massachusetts, "*Yo* he olvidado ya cómo se resuelve el cubo, pero mis *dedos* sí lo recuerdan".

Por término medio, un operador consta de 10 a 20 q-giros. Perder el hilo a medio camino de la ejecución de un operador no tiene gracia ninguna, porque en tal caso nos queda entre las manos un cubo en puro caos, aunque estuviéramos ya en la operación final. Tanto es así, que el maestro Bernie Greenberg (a quien, en colaboración con Dave Christman, tenemos que agradecer el precioso gráfico del cubo, obtenido por ordenador, que embellece la portada de este número) me dijo en una ocasión: "Si alguien me gritase "¡Fuego!" mientras estoy resolviendo un cubo, terminaría mi transformación antes de poner pies en polvorosa".

Mi propio estilo es, probablemente, de un empirismo radical. No sólo no me planteo por qué han de funcionar mis operadores mientras los estoy efectuando: tengo que admitir también que en algunos de ellos no tengo ni remota idea de por qué van bien. Descubrí estos "operadores mágicos" tras largos y arduos tanteos. Ensayé también diversas técnicas heurísticas: "tantear diversas potencias de sucesiones sencillas", "emplear conjugados liberalmente", y otras por el estilo. En cambio, apenas si me molesté en visualizar tridimensionalmente los procesos (Rubik no me lo tenga en cuenta). Conozco bien a un maestro cubero de Stanford, Jim McDonald, que puede ir explicando el motivo que inspira cada uno de los q-giros que realiza. Ninguno de sus operadores le parece "mágico", ya que es capaz de comprender cuál es su función en cada paso del proceso. Es más, ni siquiera parece sabérselos de memoria, como tuve que hacer yo con los míos; más bien parece irlos reconstruyendo conforme va reordenando el cubo, fiando siempre en su "intuición cúbica". Es como un músico experto, capaz de improvisar donde otros tienen que tocar de memoria. A los lectores interesados puede agradarles saber que la idea guía del método de Jim es resolver primero la capa superior, menos un vértice, y dejar después una columna vertical de tres cubículos que sirve de pista auxiliar de maniobras, como si fuera un camino lateral que nos permite dar fácilmente la vuelta al coche. Las otras capas se van despejando ayudándose del "apartadero" donde estacionar o recoger cubitos.

Quizá no sea coincidencia que el análisis abstracto del problema de reordenación haya sido llevado a sus últimas consecuencias por Morwen B. Thistlethwaite, compañero de despacho de Singmaster (¡no estoy bromeando!). Thistlethwaite detenta en la actualidad el record mundial de algoritmos capaces de reordenar el cubo en mínimo número de etapas. Su algoritmo nunca precisa más de 52 "giros". (Por giro entendemos bien un q-giro, bien una rotación de 180 grados en una cara.) Thistlethwaite se ha inspirado en nociones de teoría de grupos para guiar a un ordenador en la búsqueda de tipos especiales de transformaciones. Su algoritmo tiene la propiedad de no dar la impresión de ir aproximándonos a la meta hasta que nos encontramos a unos pocos giros del final.

Vale la pena contrastar esta técnica con estilos más convencionales de programación de tales algoritmos. Casi todos ellos comienzan dejando una capa totalmente terminada, la superior, por ejemplo. Al decir "capa" y no "cara" superior, quiero hacer notar que también la "cenefa" ha de quedar perfecta: los cubitos situados en lo alto han de estar correctamente alojados tanto al

ser observados desde arriba como desde los lados. Tenemos así el primero de una serie de "estados planicie". Si bien para proseguir es necesaria la destrucción temporal de estos estados, serán restaurados más adelante, y cada vez que así sucede aumenta la ordenación del cubo.

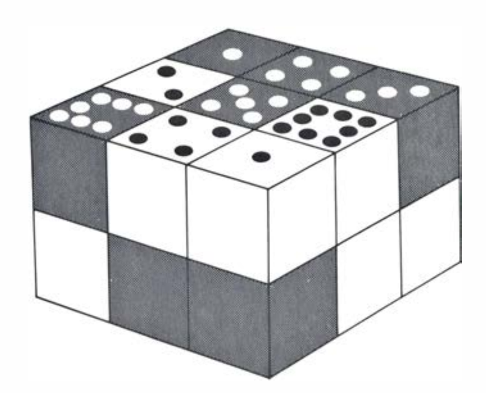
Tras dejar terminada la capa superior, el cubista acostumbra a trabajar en los vértices de la inferior, o quizás, en dejar lista la sección ecuatorial horizontal. De hecho, casi todos los algoritmos pueden descomponerse en cinco etapas naturales, correspondientes a las clases naturales de cubitos que van retornando a los cubículos que serán su definitivo hogar. Mi algoritmo particular, por ejemplo, pasa por los cinco estadios siguientes: aristas superiores, vértices superiores, vértices inferiores, aristas ecuatoriales y aristas básicas. En las dos primeras etapas consigo simultáneamente la orientación y el alojamiento correctos. Cada una de las tres últimas se descompone en una fase de alojamiento y una fase de orientación. Naturalmente, los operadores correspondientes a cada etapa han de respetar todas las conquistas de las etapas precedentes. Esto significa que sólo pueden destruir el grado de ordenación ya conseguido en tanto sean capaces de repararlo. Poco importa, empero, si revuelven y embrollan indiscriminadamente cubitos a los que prestaremos atención más adelante. He observado que, por lo común, los algoritmos de otras personas se basan más o menos en las mismas clases de cubitos, pero el orden de las etapas puede ser completamente diferente.

Prácticamente, todos los algoritmos tienen la propiedad de que si fuésemos sacando instántaneas del cubo en los estados planicie podríamos ver grupos completos de cubitos que van "entrando en razón", de acuerdo con ciertas pautas. Observada a nivel de operadores, esta propiedad muestra la "monotonía" o convergencia del sistema, perceptible en la aproximación del cubo a su estado fundamental, sin retrocesos. Evidentemente, se producen retrocesos a nivel de *giros* individuales, pero ese es otro asunto.

El algoritmo de Thistlethwaite tiene muy distinta inspiración. En lugar de intentar alojar en sus cubículos a determinadas clases de cubitos, procede por "descenso a través de una sucesión de subgrupos encajados". Significa esto empezar con total libertad, realizar unos cuantos movimientos, ceñirse después a los tipos de movimientos que serán permisibles en lo sucesivo, hacer unos cuantos movimientos más, limitar todavía más la clase de movimientos admisibles, y así proseguir hasta que las restricciones impuestas a los movimientos sean tan estrictas que ya nada se pueda mover en absoluto. ¡Justamente en ese instante se habrá conseguido la posición fundamental del cubo! En cada etapa, la restricción consiste en prohibir q-giros sobre dos caras opuestas, autorizando en lo sucesivo tan sólo medias vueltas. Las primeras caras en quedar así bloqueadas son T y B; después les toca a F y A, y finalmente, a D y a I. Lo curioso de este enfoque del problema es que no se aprecia ninguna aproximación a la posición fundamental, ni siquiera tomando la serie de fotografías en momentos cuidadosamente elegidos. Mas, de pronto, ¡zás! ya está terminado.

Lo cual suscita una pregunta bien natural: ¿existe algún procedimiento sencillo para saber cuánto falta para terminar? Es evidente que tal dato sería de gran utilidad. Por ejemplo, es bastante fastidioso tener que recurrir a toda la potencia de un algoritmo general para resolver el embrollo que un amigo haya podido crear con cuatro o cinco giros casi inocuos. Tan sólo por esta razón ya sería muy interesante poder evaluar rápidamente si un determinado estado es "verdaderamente aleatorio", o si está cercano al fundamental. Claro que, ¿cómo explicar el significado de "cercano"? Hay dos formas bastante naturales de medir distancias entre dos estados de tan vasto espacio. Por ejemplo, podríamos contar el número de q-giros, o simplemente, de giros necesarios para ir de un estado al otro (donde giro significa, como antes, bien un q-giro, bien una media vuelta). Pero sin un examen exhaustivo, ¿cómo estimar el número de giros necesarios para retornar al estado fundamental? Sería preferible una estimación fiable y razonablemente exacta, que pudiese lograrse sin más que examinar sumariamente el estado del cubo. Una sugerencia que peca de ingenua es contar el número de cubitos desplazados de sus alojamientos iniciales. Tal estimador puede dar indicaciones totalmente erróneas en estados como el llamado "lunares", donde casi todos los cubitos se encuentran fuera de sus nichos. En realidad, "lunares" sólo dista ocho q-giros del estado fundamental. Quizá pueda lograrse que el revirado y el número de quarks formen parte de estimadores más atinados, pero no conozco ninguno.

Hay un profundo y complejo razonamiento fundamentado en teoría de grupos que establece una cota superior para el máximo alejamiento de la posición fundamental: 22 o 23 giros. Este resultado es tanto más sorprendente cuando se piensa que, en sus primeros algoritmos, casi todos los cubistas precisan varios centenares de giros, y que algoritmos ya muy pulimentados requieren aún entre 80 y 90 giros. Más aún, hay muchos operadores individuales que necesitan un número de giros considerablemente mayor que el requerido por el algoritmo de Thistlethwaite completo. (Mi primer virador de doble arista tenía casi 60 giros.)


El Dominó Mágico en revoltillo

Lo que sí sabemos seguro, y podemos mostrar fácilmente, es la existencia de estados distantes al menos 17 giros del inicial. El razonamiento se desarrolla como sigue: de partida, podríamos hacer 18 giros distintos: I, I', I^2; D, D', D^2... y así para todas las demás caras. Hay a continuación 15 giros razonables posibles (pues no vamos a dar otro giro a la misma cara). El número de distintas series de giros de longitud 2 es, por tanto, 18×15, o sea, 270. Cada nuevo giro origina un nuevo factor 15. ¿Cuánto podríamos continuar antes de rebasar el número de estados accesibles? Resulta que con 17 giros se crean ya más sucesiones de giros que estados distintos pueden existir, mientras que con 16 giros ocurre lo contrario. Ahora, de ninguna manera es cierto que distintas sucesiones de giros produzcan estados diferentes y, por tanto, lo que hemos demostrado no es que con 17 giros pueda alcanzarse todo estado accesible, sino que hacen falta *al menos* 17 giros para alcanzar desde el estado inicial cada uno de los estados de su órbita. Es por ello concebible que no haya ningún par de estados distantes entre sí mucho más de 17 giros. ¿Qué 17 giros? Amigos, ésa es la cuestión.

Hoy por hoy, tan sólo el Todopoderoso sabe cómo pasar de uno a otro estado del cubo mediante un número mínimo de giros. El "algoritmo divino" sería, pues, el que explicase cómo con-

seguirlo. Una cuestión candente de la ciencia cubológica es precisamente: ¿será el algoritmo divino una tabla gigantesca sin pautas discernibles, o bien habrá en él la regularidad suficiente como para construir un algoritmo breve y elegante, que los meros mortales podamos dominar?.

De todas formas, si Dios tuviera que tomar parte en un concurso de resolución del cubo encontraría serios contrincantes en algunos mortales prodigiosos, a pesar de no disponer éstos de Su algoritmo. ¡Me han contado que un joven inglés, Nicholas Hammond, de Nottingham, ha conseguido rebajar el tiempo medio de resolución a cerca de 30 segundos! Semejante proeza requiere cierto número de habilidades. La primera es una profunda comprensión del cubo. La segunda, un sistema de operadores extremadamente pulido. La tercera, tener los operadores tan "integrados" en uno mismo que se sepan realizar hasta en sueños. La cuarta es pura velocidad en el juego de muñeca. Y la quinta, disponer de un "cubo de carreras" bien aceitado; de un cubo que gire a la más ligera indicación de los dedos, que adivine ansioso cada operador casi antes de ser requerido. En resumen, un cubo de carreras es un cubo que *desea* ganar.

No sé todavía de nadie que haya bautizado a sus cubos de carreras, aunque seguro que terminará por ocurrir. Por otra parte, parece como si para contribuir a la ciencia del cubo fuera necesario tener apellidos curiosos. Además de Singmaster y Thistlethwaite, tenemos a Dame Kathleen Ollerenshaw (quien recientemente recibió el título de Lord Mayor de Manchester). Dame Kathleen ha descubierto muchos airosos procesos, escrito un artículo sobre el Cubo Mágico y tenido la distinción de sufrir el primer ataque agudo confirmado de "pulgar de cubero", forma grave de la enfermedad mencionada al comienzo de este artículo. Tenemos después a Oliver Pretzel, descubridor de un precioso 3-ciclo de torsión y creador de un precioso "motivo" llamado "estado 6-T", alcanzable desde el estado fundamental con ayuda de la voz $I'D^2F'I'A'TAIFDT'DID_sF_sT_sD_s$.

Desde siempre, los "motivos curiosos" han interesado a los cuberos aficionados, pero no podré hacerles justicia. Sólo mencionaré algunos de los mejores. Mi favorito es el "Gusano", cuyo "genotipo" o sucesión de giros es $DTF^2B'D_sF_sB'F'D'F^2DT^2FD^2F'D'T'F'T^2FD$. *Tenemos después la "Sierpe"*, motivo sinuoso parecido al gusano, que se arrolla reptando alrededor del cubo: $AD_sB'D^2BD'_sA'D^2TA^2T'BD^2B'$. Amputándole la cola a la Sierpe (D^2B'), e injertándole en cambio $A^2D_aT^2D'_aA^2B'$ se crea un nuevo motivo, dos aros entrelazados, muy curioso. Todas las configuraciones anteriores se deben a un maestro en tales diseños, Richard Walker.

Entre los motivos más hermosos tenemos el Mesón Gigante, que se construye a partir de un quark gigante (un subcubo de 2 × 2 × 2 girado 120 grados) y de un antiquark gigante. Para rizar el rizo podemos usar quarkilladores y formar un quark y un antiquark de tamaños normales sobre los vértices del quark y antiquark gigantes, como la guinda que corona un merengue [*véase la ilustración de la página 123*]. Dejaré a los lectores la tarea de lograrlo. Como ejercicio de precalentamiento puede ir bien la configuración llamada Pons Asinorum (Puente de burros), así llamado porque, como me hizo notar uno de los cubomaestros del MIT, "si no se consigue montar este pollino, más vale que vaya pensando en dejar la cubo-hípica". Tenemos ahora dos clases de cruces, que en la escuela superior de ecubitación del MIT se conocen por "cruz de Plummer" y "cruz de Christman" [*véase la ilustración citada*]. En la cruz de Christman se usan tres pares de colores (*T-B, F-D, I-A* en la ilustración); la cruz de Plummer requiere dos ternas del tipo quark-antiquark.

Me gustaría dejar al lector una serie de indicaciones y cuestiones en que reflexionar. Un difícil reto, válido para cubistas de todos los niveles, consiste en hacer que otra persona le dé a un cubo prístino unos cuantos giros; la tarea consiste en tomar este cubo ligeramente embrollado y devolverlo a la posición fundamental, hallando la palabra inversa *mentalmente.* Los expertos podrán lograrlo con movimientos más complejos que los noveles. Parece que Kate Fried es normalmente capaz de atinar siete giros, y que en una ocasión, tras estar todo un día con la vista fija en el cubo, ¡pudo invertir 10 giros! Por mi parte, yo lo consigo con unos cuatro.

Mi camino real para descubrir algoritmos se basa en dos estimulantes ejercicios que sólo hacen uso de cubitos triédricos. El ejercicio preliminar es como sigue: maniobre con los cuatro vértices que porten una faceta blanca hasta llevarlos a la tapa, con las caras blancas hacia lo alto, y sin prestar atención a qué cubos ocupan qué cubículos. Haga simultáneamente otro tanto con la base (naturalmente, con *su* color señalando hacia abajo). Los cubistas aventajados han de realizar el mismo ejercicio, ase-

gurándose además de que todos los vértices queden en los cubículos correctos. El problema equivale a resolver un cubo mágico de 2 × 2 × 2; quien lo consiga habrá recorrido buena parte del trecho que conduce hasta el magisterio en cubos mágicos.

Para facilitarle los procesos de permutación de aristas, he aquí un maravilloso truco descubierto por David Seal, basado en un tipo de operador llamado monoflip, o sea, monovirador. Se lo presentaré como problema. ¿Cómo se puede construir un doble virador de aristas basándose en un proceso que embrolla las dos capas inferiores, pero que deja invariante la capa superior con excepción de un solo cubito-arista, que sufre virado? Les daré una pista: en la solución se emplea un importante concepto de teoría de grupos, la noción de conmutador, palabras de la forma *PQP'Q'*. Dejaré también como ejercicio la tarea de hallar operadores monoflip. Cuando yo lo hube descubierto, inmediatamente hice sitio en mi método para darles cabida.

He aquí un pequeño acertijo: ¿por qué causa se producen con tanta frecuencia ciclos de longitudes 5 y 7 en un objeto cuyas simetrías tienen todas relación con números como 3, 4, 6 y 8? ¿Qué origina entonces las longitudes 5 y 7? Otra cuestión no muy dispar: ¿cuál es el máximo orden que puede tener una palabra? Se llama orden al mínimo exponente a que debemos elevarla para que la potencia sea la identidad. Por ejemplo, *D* tiene orden 4. Inspeccionando su expresión como producto de ciclos, podemos llegar a que *DT* tiene orden 105.

¿Hacia dónde van los nuevos desarrollos? Ante todo, en esta reseña apenas si he rozado la superficie de la cubología. Rubik y otros están ya trabajando en diversas generalizaciones. Existe un Dominó Mágico, algo así como dos tercios de un Cubo Mágico, con dos caras de 3 × 3. Sólo es factible hacerlo girar medias vueltas en torno a dos de sus tres ejes, admitiendo *q*-giros en torno al tercero. En la posición fundamental una cara es negra y, la otra, blanca; ambas portan en orden los números del 1 al 9. Aún se asemeja al rompecabezas de Loyd más que el propio Cubo Mágico. Algunas personas han construido cubos de 2 × 2 × 2, que quizás algún día salgan a la venta. Podemos fabricarnos uno nosotros mismos, pegando pequeños sombreros tricornios sobre cada uno de los ocho vértices de un cubo de 3 × 3 × 3. Es evidente que los lectores se habrán preguntado sobre la posibilidad de cubos de 4 × 4 × 4. Por si les tranquiliza, están poniéndolos a punto en Holanda; quizá queden listos muy pronto. Es también inevitable plantearse el problema en dimensiones mayores o menores. Y los teóricos están comenzando ya a examinar las propiedades de cubos de dimensión superior.

Mas el potencial del cubo de 3 × 3 × × 3 dista mucho de quedar agotado. Un fértil terreno, prácticamente inexplorado, es el de los coloreados alternantes. Tal idea me fue sugerida por varios jilmaestres del MIT. Los cubitos pueden pintarse de diversas formas; cada coloreado nuevo presenta un problema de reordenación diferente; así, en uno de ellos las orientaciones de los cubitos-arista carecen de importancia y en cambio resulta crucial la de los centros de las caras. En otro, son los vértices los irrelevantes, mientras los centros adquieren máxima importancia. Si deseamos simplificar las cosas, podemos pintar dos caras del mismo color, reduciendo el número de éstos de seis a cinco. O bien podemos usar sólo tres colores para las caras. En el límite, podríamos tener tres caras blancas concurrentes en el vértice y tres caras azules convergiendo en el diagonalmente opuesto. El funcionario francés que mencioné al principio dice que en los cubos que vio, cinco caras eran de un color, ¡y la sexta de otro!

¿Quién podrá decir en qué terminará todo? Ya lo ha hecho notar Bernie Greenberg: "El cubismo exige del cubóforo la invención de una ciencia, en sentido estricto. Cada uno debe proponerse campos de investigación, diseñar experimentos, descubrir principios generales, construir teorías, verificarlas, etcétera. Sin duda, es el único rompecabezas que exige de sus resolventes la construcción de toda una ciencia". ¿Podrían Rubik e Ishige haber soñado que sus inventos iban a crear un modelo, una metáfora, de todo lo profundo y hermoso de la ciencia? Nadie pondrá en entredicho que sus cubos son objetos admirables

Quisiera aprovechar la ocasión para agradecer a tantos cubistas, cubólogos y cubomaestres su ayuda, colaboración y estímulo. Entre ellos están Clark Baker, Alan Bawder, Jim Boyce, Larry Breed, Charles Brenner, Bob Filman, Carl Hoffman, Scott Kim, Bill McKeeman, Jeanine Moseley, Richard Pavelle, Dave Plummer, David Policansky, David Singmaster, Ann Trail, Allen Wechsler, Dan Weise y John Woodcock. Mi agradecimiento es máximo a Greenberg, quien organizó una sesión cubológica en el MIT a beneficio mío, y que además me contagió de su entusiasmo por la cubología.

Taller y laboratorio

Movimiento de las peonzas, con referencia a algunos tipos excepcionalmente originales

Jearl Walker

Aunque la peonza es un juguete de orígenes muy remotos, los principios que rigen su comportamiento no empezaron a conocerse hasta hace siglo y medio. ¿Por qué un objeto de formas irregulares puede mantenerse girando sobre un solo punto? ¿Por qué peonzas de formas diferentes se comportan de modos tan distintos? Voy a ocuparme de la mecánica de las peonzas, pero evitando el follaje matemático que suele ocultar el tema en algunos textos de física. Presentaré, además, algunas de las insólitas peonzas que construye Donald W. Dubois, de la Universidad de Nuevo México, quien ha puesto a bailar, cual trompos, objetos tan dispares como un tee de golf y el tapón de una botella de tinta china.

El fundamento del comportamiento de las peonzas se encuentra en la acción de la gravedad. Esta atrae hacia abajo todos los átomos de una peonza; pero la atracción resultante se imagina mejor suponiéndola aplicada al centro de masa del trompo, situado en un punto de su interior, usualmente en su centro geométrico. El peso de la peonza puede representarse mediante un vector aplicado al centro de masa y dirigido hacia abajo. Parece lógico entonces que, a causa del peso, haya de ser más proclive a volcarse (como nos ocurre a nosotros cuando nos inclinamos respecto a la vertical) que a mantenerse erguida.

Pero hay una diferencia, y es que la peonza gira sobre sí misma, en virtud de lo cual la atracción gravitatoria origina la sorprendente rotación del juguete alrededor de la vertical. No es fácil imaginar esta reacción, habituados como estamos a los objetos fijos o en movimiento lineal. En general, una fuerza aplicada a un objeto produce en éste una aceleración en la dirección de la fuerza. Pero si interviene algún movimiento de giro, la fuerza puede desencadenar un movimiento perpendicular a ella misma. Este modo de moverse particular forma parte del encanto de las peonzas.

Los trompos poseen momento cinético. Se entiende por tal el producto de su momento de inercia (que da cuenta de cómo está distribuida su masa) por su velocidad angular; la cantidad de movimiento de un cuerpo es el producto de su masa por su velocidad lineal. Una peonza posee un momento cinético porque gira alrededor de su eje de figura y porque posee una masa distribuida en torno a dicho eje. El momento cinético es un vector situado sobre el eje de figura, que es el eje de simetría del trompo.

El único modo de alterar el momento cinético de un cuerpo es mediante un momento de giro. Científicamente, el momento de giro se define como el producto de (1) la fuerza ejercida sobre el cuerpo por (2) el brazo de palanca, que es la distancia en perpendicular que separa el centro de giro de la recta que contiene la fuerza. En el caso de una peonza, el centro de giro se encuentra en su extremo puntiagudo que establece contacto con el suelo o con el tablero de una mesa. El brazo de palanca se extiende horizontalmente desde dicho punto hasta la vertical que pasa por el centro de masa, que es donde se considera que la gravedad ejerce su atracción sobre la peonza.

Pero la gravedad no produce solamente una fuerza dirigida hacia abajo, sino también un momento de giro capaz de alterar el momento cinético de la peonza. Y lo hace de un modo sencillo: rectificando la dirección del vector momento cinético y haciendo que éste dé vueltas alrededor de la vertical. Por tanto, como el momento cinético debe seguir estando sobre el eje de figura de la peonza, éste también dará vueltas alrededor de la vertical ejecutando el movimiento llamado precesión y que describe un cono centrado en la vertical. (Voy a suponer, por ahora, la punta de la peonza como si estuviera fija en el punto donde toca el suelo.)

Si la peonza, vista desde arriba, gira en sentido antihorario, la precesión en torno a la vertical será también antihoraria. Cuando el giro sea horario, la precesión será horaria. Si la peonza no estuviese girando, la acción del momento de giro creado por la gravedad haría que la peonza cayera sobre el suelo, como realmente ocurre durante la última etapa del baile en virtud del efecto del rozamiento sobre el punto sobre el que gira el trompo. En la terminología de los juegos de peonza, esto último se expresa diciendo que la peonza se muere; mientras está bailando se dice que la peonza está viva.

Casi todas las peonzas se inclinan un poco en cuanto empiezan a bailar, efecto que obedece a un requerimiento energético. Para que la peonza precesione, en razón del momento de giro creado por la gravedad, debe poseer energía cinética suficiente. Aunque puede adquirir esa energía en el lanzamiento, lo común es que proceda de la inclinación inicial, de manera que la disminución de energía potencial de la peonza se transforme en energía cinética.

Esta explicación tan elemental se derrumba la primera vez que nos disponemos a examinar el movimiento de una peonza viva. Se observa entonces que el trompo no se limita a girar y precesionar, sino que además su eje de figura cabecea ejecutando la llamada nutación. Al tiempo que nuta, el ángulo formado por su eje de figura y la vertical varía entre dos valores determinados por la distribución de masa y la energía cinética inicial de la peonza y el ángulo inicial de ésta con la vertical. Los tipos de nutación pueden ilustrarse siguiendo la posición de la intersección del extremo superior del eje de figura con una esfera cuyo centro sea el punto de contacto del juguete con el suelo. Se vería así que el ángulo formado por el eje de figura de la peonza y la vertical está limitado por dos círculos horizontales centrados en la vertical y que representan los límites entre los cuales el trompo puede inclinarse, respecto a la vertical, durante la nutación.

En uno de los tipos de nutación, el eje de figura zigzaguea armónicamente entre los dos círculos límites, rozándolos tangencialmente. El sentido de precesión del eje es, en ese caso, siempre el mismo, horario o antihorario vista la peonza desde arriba. En la ilustración de la página 141 la rotación es antihoraria.

En un segundo tipo de nutación, el eje de figura riza el rizo entre los dos

circulos límites, pero sigue siendo tangente a ambos. El sentido del movimiento del eje cambia aquí, periódicamente, de horario a antihorario, pese a lo cual el valor medio de la precesión corresponde a un sentido o al otro. En la ilustración citada, la precesión media es antihoraria.

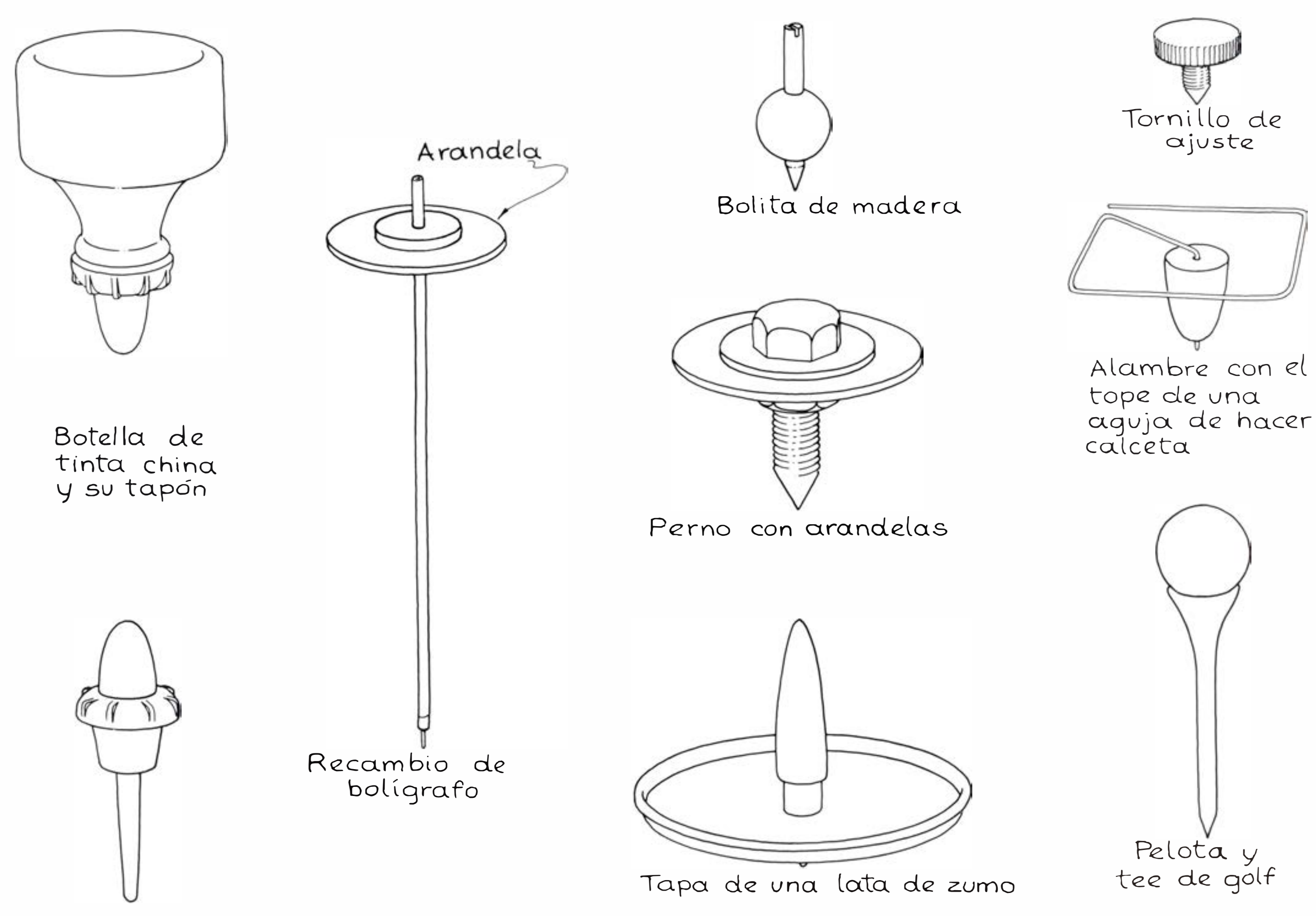
En el tercer tipo de nutación, se trazan guirnaldas sobre la esfera imaginaria, de suerte que la trayectoria del eje de figura toca tangencialmente a uno de los círculos, como ocurre en los otros dos casos, y perpendicularmente al otro círculo. La precesión tiene siempre el mismo sentido, pero la velocidad de precesión varía desde un máximo en el círculo inferior hasta cero en el superior.

Si se presenta, el tipo de nutación dependerá de las condiciones iniciales del giro de la peonza. Sobre estas condiciones nunca se posee un dominio completo, dadas las numerosas perturbaciones que se producen al lanzar el juguete. Supongamos que el lanzamiento imprime a la peonza una velocidad de precesión de igual sentido al de la velocidad de precesión debida a la gravedad. La peonza precesionará entonces con nutación armónica e, independientemente de cuál sea su posición durante su ciclo de nutación, tanto su velocidad de precesión inicial como la velocidad debida a la gravedad garantizarán una precesión en torno a la vertical cuyo sentido será ininterrumpidamente el mismo.

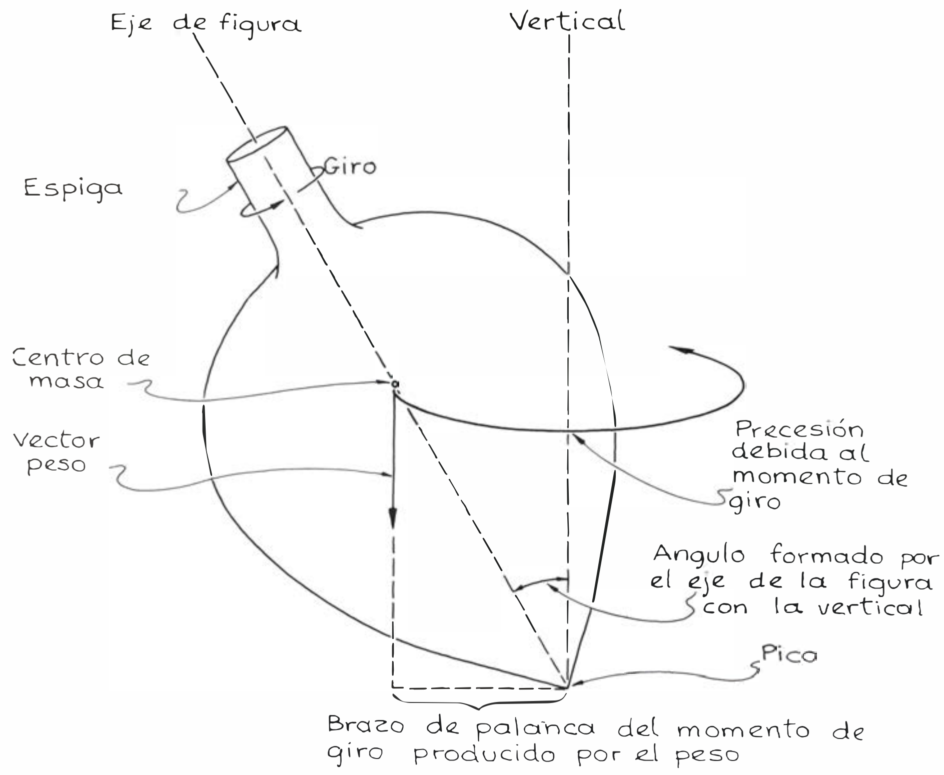
Si la velocidad de precesión inicial es opuesta a la precesión debida a la gravedad, la nutación será rizada. Aquí, mientras se describe la parte inferior del rizo, la precesión gravitatoria impulsa a la peonza en torno a la vertical según el sentido de preferencia. Mientras se describe la parte superior del rizo, que es cuando el eje de figura alcanza el círculo límite superior, la precesión producida por la gravedad se agota y sólo queda la precesión inicial. Entonces, hasta que la peonza descienda y vuelva a ser impulsada por la gravedad, precesionará en el sentido contrario.

El tercer tipo de nutación suele aparecer cuando se sujeta inicialmente la peonza formando un ángulo no nulo con la vertical y se suelta luego. Al carecer de precesión inicial, la peonza cae inmediatamente sobre el círculo límite inferior. Esta caída es necesaria porque la disminución consiguiente de energía potencial de la peonza facilita la energía para la precesión producida por el momento de giro ejercido por la gravedad. Cuando la nutación continuada devuelve el eje de figura a su ángulo inicial con la vertical, la peonza recupera su energía potencial inicial y la precesión se detiene momentáneamente, ya que no queda energía para ella. Tras lo cual, la peonza vuelve a caer sobre el círculo inferior y la precesión prosigue hasta que la nutación vuelve a elevarla.

El origen de los círculos límites se halla en tres restricciones rigurosas que sufren los movimientos de giro, precesión y nutación. Primero, la energía total (cinética más potencial) de la peonza debe permanecer constante. (De los efectos del rozamiento me ocuparé más adelante.) Segundo, el momento cinético en la dirección del eje de figura debe permanecer constante, aunque puede cambiar de dirección, ya que no existe ningún momento de giro en la dirección de dicho eje que pueda hacerlo variar. Finalmente, y por la misma causa, el momento cinético según la vertical debe permanecer constante. (Una vez comenzada la precesión, el cálculo de



Objetos convertidos en peonzas por Donald W. Dubois



Momento de giro actuante sobre la peonza

esos dos momentos cinéticos se hace más arduo, porque la velocidad angular incluirá entonces el giro de la peonza y su velocidad de precesión.) A causa de estas tres condiciones estrictas, el ángulo que forma el eje de figura con la vertical se ve forzado a variar entre unos márgenes determinados. Y si el eje se inclinara por debajo del círculo límite inferior, o se elevara por encima del superior, lo haría incumpliendo tales condiciones. (En realidad, las peonzas sí se elevan por encima del círculo superior, pero únicamente a causa del rozamiento.)

La mayoría de los modelos matemáticos que se han ideado para la peonza se basan en la hipótesis simplificadora de que la energía cinética de rotación es muy superior a la variación de energía potencial producida por el cabeceo de nutación. De una peonza como esta se dice que es "rápida". Sentada la hipótesis, hay algunas características del movimiento que pueden relacionarse con la velocidad de rotación de la peonza sobre sí misma. Si esta velocidad es menor, aumenta el cabeceo de nutación pero disminuye la velocidad de nutación. La velocidad de precesión depende también del giro, siendo mayor cuando más lento sea éste. Estas relaciones pueden apreciarse fácilmente cuando se hace bailar una peonza sobre el suelo. A medida que el rozamiento resta vida a la peonza, aumenta la velocidad de precesión y la nutación se hace más lenta y más pronunciada. Finalmente, instantes antes de morir, la peonza se balancea pesadamente arriba y abajo mientras precesiona en torno a la vertical con más rapidez que nunca.

Variación del momento cinético producida por el momento de giro
Sentido de la precesión
Momento cinético en un instante determinado
Nueva dirección del momento cinético

Efecto del momento de giro

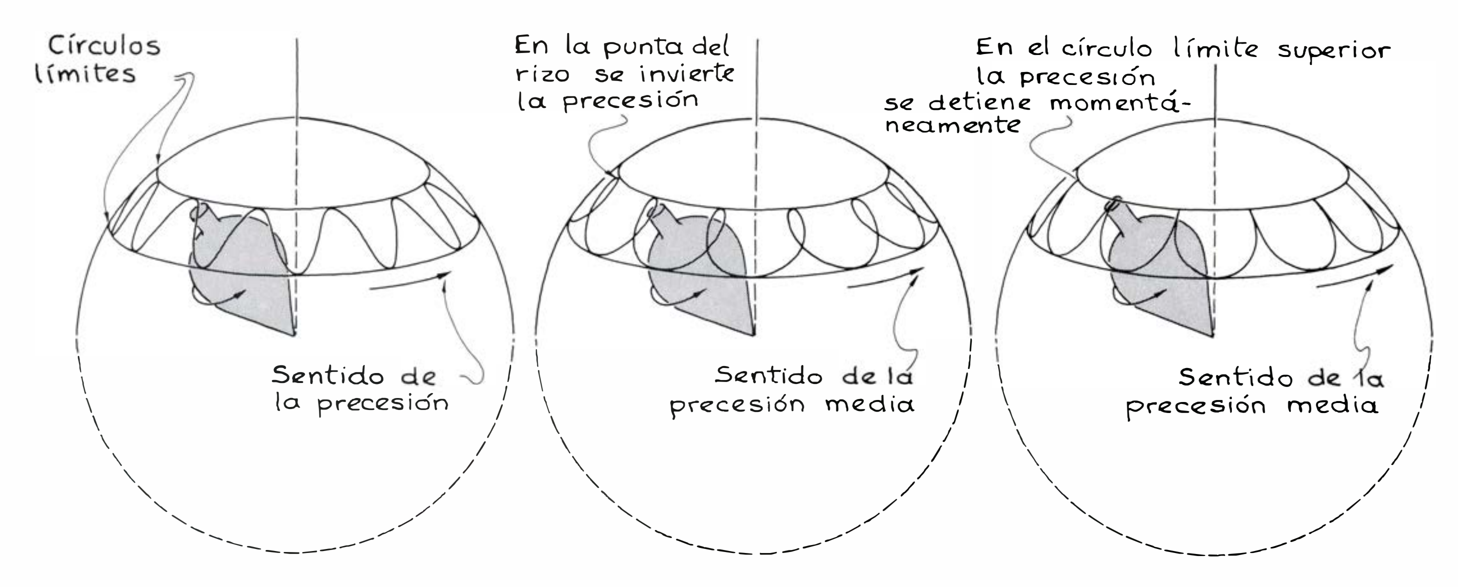
La velocidad de precesión y la amplitud y frecuencia de la nutación dependen también de la masa y la forma de la peonza. En general, una masa mayor incrementa ambas componentes del movimiento. (Desde luego, cuesta más poner a bailar a una velocidad dada una peonza de mayor masa que otra de menor masa.)

Si la peonza gira muy rápidamente, la poca nutación que pudiera presentarse podría resultar eliminada por el rozamiento sobre la punta de la peonza. La peonza parecería precesionar entonces, uniformemente y sin nutación, en torno a la vertical. Sin embargo, esta regularidad se debe sólo a la intervención del rozamiento.

Cuando los círculos límites son coincidentes, puede darse una precesión exactamente uniforme. Para que así sea, hay que lanzar el trompo de modo que el ángulo que forme con la vertical corresponda con el ángulo de los círculos límites. Una peonza así no nuta, al no poder elevarse por encima de los círculos ni descender por debajo de ellos. Al no descender, tampoco puede convertir parte de su energía potencial en la energía necesaria para la precesión; será, pues, el procedimiento de puesta en movimiento o el impacto inicial sobre el suelo lo que deberá producir el impulso necesario para iniciar la precesión.

En realidad existen dos tipos de precesión posibles. De ellos, sólo el más lento está relacionado con la atracción gravitatoria sobre la peonza. Aunque el rápido se observa rara vez, puede presentarse si las condiciones iniciales son adecuadas. En la explicación clásica de la precesión rápida interviene una peonza (que en las exposiciones de los textos se considera habitualmente un elipsoide) que puede rodar libremente sobre un plano horizontal, a la vez que gira sobre sí misma. Sobre esta peonza imaginaria no actúan fuerzas ni momentos de giro, ni siquiera la gravedad. En esa situación, el trompo no puede resbalar sobre el plano para caer en él, ni romper su contacto con el mismo.

Mientras baila la peonza con su eje inclinado respecto a la vertical, comienza a precesionar en torno a ésta, de suerte que mantiene contacto con el plano horizontal, que recibe el nombre de plano invariante; la trayectoria des-



Formas de nutación de una peonza

crita sobre el mismo por la punta de la peonza se llama herpolodia y, polodia, la trayectoria descrita por la punta de la peonza sobre la misma peonza. El movimiento así descrito conduce a una afirmación que, como ya dijo alguien, parece un galimatías: "la polodia rueda sin deslizar sobre la herpolodia que yace en el plano invariante". Esta precesión constituye una idealización de la precesión rápida que podría tener un trompo de verdad.

Mi colega James A. Lock ha puntualizado que la precesión rápida puede observarse en el giro de una pelota de fútbol americano y en una botella de refrescos lanzada al aire. Los jugadores de la línea media lanzan las pelotas dotándolas de giro al objeto de que mantengan la estabilidad en vuelo, aprovechando su forma aerodinámica. En los pases largos, las pelotas suelen llevar un notable giro excéntrico, que en parte puede deberse a la resistencia del aire. Incluso sin esta resistencia, la pelota giraría excéntricamente a condición de que se lanzara de modo que desarrolle una precesión rápida.

Esta precesión se observa mejor si se lanza hacia arriba una botella de refrescos dotada de giro. (La botella debe estar vacía, pues el bailoteo del líquido interferiría con el giro.) Si la botella se lanza con el eje vertical, quizá no se observe precesión alguna. Lanzándola con menos cuidado, la botella, a la vez que gira sobre sí misma, precesionará en torno a la vertical en el modo rápido. Este movimiento no se debe a la gravedad, pues la botella está volando. Durante la precesión, el fondo de ésta describe la herpolodia sobre un plano invariante perpendicular a la vertical.

La precesión rápida apenas aparece en el juego, porque es mucho más rápida que la precesión dependiente de la gravedad. Para dotar a una peonza de precesión rápida se necesita mucha más energía y, a menos que la reciba de algún modo en las etapas iniciales (por ejemplo, si fortuitamente la recibe en el lanzamiento acompañada del movimiento inicial conveniente), en lugar de ello se establecerá una precesión lenta.

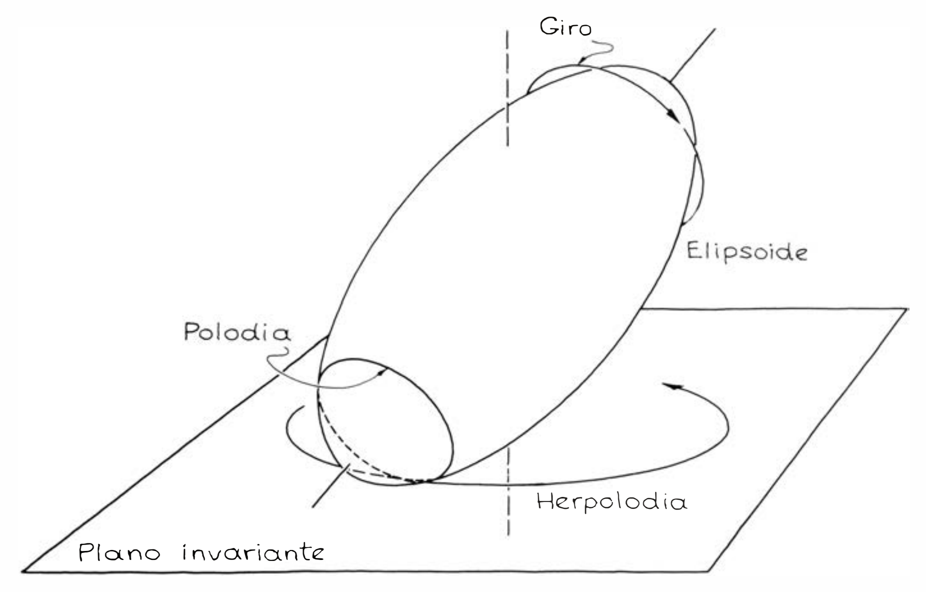
Con tino y cuidado podemos poner a bailar una peonza de modo que su eje quede vertical. (Aquí es muy útil el procedimiento de lanzarlo con dos cabos de cuerda que les describiré después.) Si la velocidad de giro sobrepasa determinado calor crítico, que depende del tipo de peonza, ésta permanece derecha. Para una velocidad de giro menor que la crítica, el juguete nuta entre la vertical y un círculo límite correspondiente a un ángulo con la vertical no nulo.

Ambos modos de comportamiento pueden observarse en una peonza real cuya velocidad de giro sea inicialmente superior al valor crítico y cuyo eje de figura esté inicialmente vertical. El giro atraviesa entonces dos etapas: primero un giro vertical sosegado y luego una nutación frenética. Este cambio tiene su causa en el rozamiento sobre el punto donde la peonza toca al suelo. Ocurre que, aun cuando la velocidad de giro inicial sea suficientemente elevada para que el giro sea vertical, el rozamiento acaba por reducir la velocidad de giro hasta un valor inferior al crítico; la peonza comienza entonces a inclinarse y a nutar y termina por morirse.

¿Puede suceder que un trompo inclinado inicialmente nute hasta la vertical? En principio no, porque siempre le faltará la energía necesaria para adquirir una verticalidad exacta. Sin embargo, puede acercarse mucho a ello si el círculo límite superior es muy pequeño. La peonza puede nutar entonces hacia arriba, de suerte que pase muy cerca de la vertical antes de volver a nutar hacia abajo. Se trata de una situación que, quizá, no se da con frecuencia.

Un observador avezado en esos comportamientos no estaría de acuerdo con casi nada de cuanto acabo de exponer. En realidad, sí hay peonzas que se levantan hasta girar verticalmente (se llaman peonzas dormidas), por lo que mis afirmaciones no son del todo correctas. Pero no se elevan a través de un efecto de nutación, sino a través de una sutil e imperceptible acción recíproca entre el pico de la peonza y el rozamiento entre la punta de la peonza (extremo del pico sobre el que gira) y el suelo. El pico de una peonza real no es el punto infinitesimal que hasta aquí he venido suponiendo; se acerca más al pico semiesférico.

Un pico como ése tendería a dar vueltas sobre el suelo en virtud del propio giro de la peonza. Además, el trompo trata de precesionar al mismo tiempo. Como ahora consideramos un pico más ajustado a la realidad, en vez de un punto inmóvil en contacto con el suelo, hemos de imaginar la precesión como un movimiento de la peonza en torno a su centro de masa, que permanece inmóvil. Vistas las cosas así, el punto de contacto con el suelo se moverá describiendo círculos por debajo del centro de masa, mientras que la parte superior de la peonza describirá círculos por encima del centro de masa.



Movimientos que engendran la polodia y la herpolodia

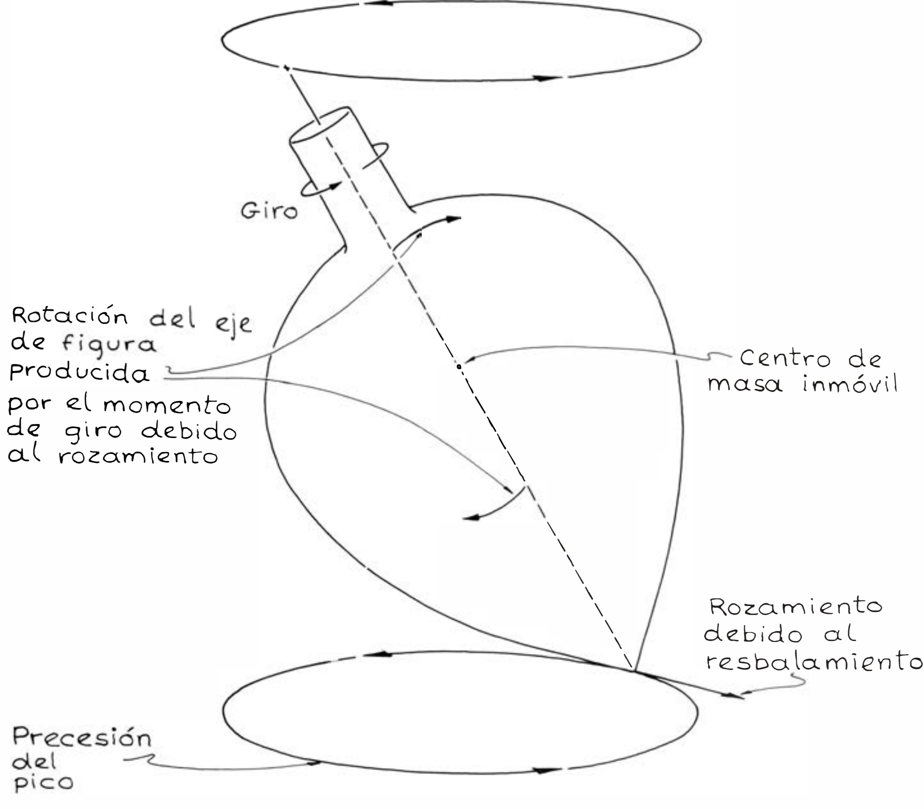
Como resultado, el pico semiesférico tratará de rodar sobre el suelo de dos modos, uno instado por la rotación de la peonza alrededor de su eje de figura y, el otro, provocado por la precesión que lo impulsa sobre el suelo. Por ser el primero más rápido que el segundo, el pico resbala. En este resbalamiento se produce un rozamiento entre el pico y el suelo de sentido contrario al resbalamiento, tal como se ve en la ilustración inferior de esta misma página; dicho rozamiento produce una rotación del eje de figura poniéndolo vertical. Para calcular el momento de giro correspondiente, podemos considerar que el centro de giro es el centro de masa y, entonces, el brazo de palanca será la distancia entre ese punto y el punto donde el pico toca al suelo; así, el momento de giro que endereza la peonza es el producto de este brazo de palanca por el rozamiento.

También es cierto que el rozamiento sobre el pico disipa energía. Por eso, el truco para diseñar una buena peonza dormida consiste en dar al pico una forma tal que el trompo se enderece antes de perder demasiada energía; por esta razón, las buenas peonzas dormidas tienen unos picos de pequeño radio de curvatura. Las peonzas con picos de radio de curvatura mayor necesitan más tiempo para enderezarse, y quizá no lleguen a la vertical antes de que la velocidad de rotación descienda por debajo del valor crítico necesario para mantener una posición vertical. Las peonzas de látigo, que se lanzan con un cordel para dar más energía al giro, muestran unos picos de radio de curvatura muy pequeño y se duermen a los pocos segundos del lanzamiento.

La inmensa mayoría de las peonzas que se venden en las jugueterías tienen forma de pera. Algunas se ponen en movimiento con las yemas de los dedos. Las más comunes entre nosotros son las que se lanzan al suelo al tiempo que se tira con energía del cordel arrollado. En ambos casos, caen girando sobre el suelo.

No se agota con ellos, ni mucho menos, la gran variedad de modelos que pueden (o que en su día pudieron) encontrarse en todo el mundo. En la bibliografía de este número se cita la obra de D. W. Gould, que constituye una hermosa descripción de los muchos tipos de peonza construidos hasta ahora. Hay peonzas dotadas de un conjunto de caras planas que sirven para jugar a suertes o para adivinar el futuro. Algunas son discos planos atravesados por un vástago delgado; las hay que son discos planos con una protuberancia central sobre la que bailan (pueden desplegarse varias de ellas de una vez desde un sólo soporte). Aparte de la peculiaridad de su movimiento, la característica más extraordinaria de las peonzas radica en que fueron desarrolladas bajo formas muy variadas por gentes que trabajaron casi independientemente en distintos puntos del mundo.

Dubois me ha escrito recientemente acerca de las peonzas que él y sus alumnos construyen a partir de objetos tan comunes como botones, agujas y tornillos. Lo corriente es que objetos tan inverosímiles necesiten sólo un pequeño anclaje para que puedan remedar a los trompos. Quizás haya que limar una punta o practicar una ranura en un eje para insertar el cordel. Otras veces unen varias piezas con pegamento para formar una peonza. (Según explica Dubois, nada da mejores resultados que la epoxia.)



Enderezamiento de una peonza a causa del rozamiento

Habida cuenta del papel que desempeña el rozamiento en el comportamiento de estos juguetes, Dubois experimenta con superficies diferentes. Comprobó que el papel, las hojas de plástico delgadas (como los envoltorios de alimentos) y las sábanas de cama dan gran resultado con algunas peonzas, pero ninguno con otras. Para Dubois, el suelo de linóleo es la única superficie sobre la que todas sus peonzas se comportan suficientemente bien, aunque casi todas ellas se comportan mejor sobre otra superficie. Si se quiere seguir la trayectoria descrita por la punta de una peonza, podemos ponerla a bailar sobre papel carbón. papel recubierto de hollín, vidrio recubierto de tinta o cualquier otra superficie donde la punta pueda dejar huella.

Algunas de las peonzas de Dubois pueden ponerse en marcha con un rápido movimiento de dedos. Sujeta la espiga de la peonza entre el pulgar y el dedo corazón, manteniendo hacia abajo la palma de la mano y la punta de la peonza; luego chasquea ambos dedos haciendo que la peonza salga girando por la mesa o el suelo. Si la peonza carece de espiga, apoya el pico sobre la palma de la mano, abierta hacia arriba; cuando pone a bailar la peonza con un brusco movimiento de dedos, la punta ya está convenientemente apuntada hacia abajo. Puede lanzarse de otra manera: se sujeta el pico entre el índice de la mano izquierda y el pulgar de la mano derecha, se retiran luego los dedos muy rápidamente en direcciones opuestas y la peonza se pone a bailar.

Con algunas peonzas pueden conseguirse rotaciones muy rápidas arrollando un cordel o una tira de tela alrededor del tronco o de cualquier otra porción cilíndrica de la peonza. Las cintas para máquina de escribir dan buenos resultados. (Ni que decir tiene que hay que eliminar antes la tinta.) Se introduce uno de los extremos de la cinta en la ranura del tronco y luego se arrolla dándole varias vueltas alrededor y se tira luego el trompo paralelamente al suelo, a la vez que se sujeta el extremo contrario de la cinta; el juguete comienza a girar entonces con inusitada rapidez.

Este procedimiento no sirve para el caso de peonzas pequeñas o frágiles. Con éstas, Dubois prefiere un arrollamiento de cinta doble. Las dos cintas se apuntan de la misma forma en la ranura y luego la peonza se hace girar sobre la mano para formar una capa de cinta doble de seis a 12 vueltas. En vez de lanzar la peonza para desenrollar las cintas, se tira ahora de éstas horizontalmente en dirección opuesta, consiguiendo que la peonza se ponga a girar. Con arrollamientos dobles de hasta 14 vueltas, Dubois hacer rodar una tachuela de tapicero a unas 15.000 revoluciones por minuto. Con los dedos logra velocidades de hasta 8000 revoluciones por minuto.

Si el tronco de la peonza es muy terso y pulido, como es el caso de las tachuelas de tapicero, cuesta mucho arrollar la cinta, porque resbala. Entonces, para aumentar el rozamiento, Dubois recubre la superficie con una capa de epoxia. Este rozamiento adicional significa que, cuando en el lanzamiento se llegue al extremo de la cinta, la peonza será desviada por un tirón final en vez de escurrirse libremente.

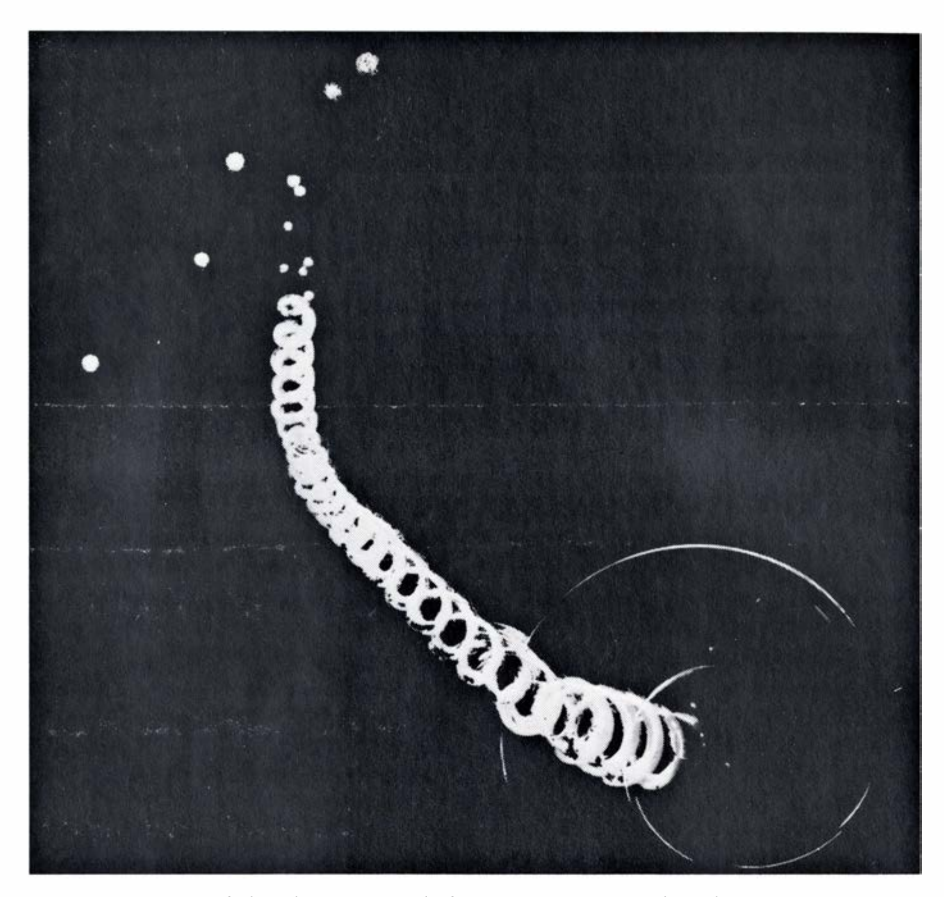
Algunas de las peonzas empleadas por Dubois silban o zumban cuando están vivas, sea porque arañan la superficie sobre la que bailan o porque generan turbulencias y vibraciones del aire. La velocidad de giro de una peonza silbante debe sobrepasar las 3300 r.p.m. Uno de sus ejemplares más rápidos silba a dos octavas por encima del la (440 hertz). Si la peonza baila sobre cierta clase de membrana, ésta actúa de tornavoz, realzando la audibilidad del sonido. La membrana no debe tensarse excesivamente ya que la peonza podría perforarla.

La rotación y la precesión pueden observarse con una lámpara estroboscópica. Para ello, la frecuencia de ésta se modifica hasta acompasarla con la rotación o con la precesión. Así se ve que algunas de las diminutas peonzas que Dubois me ha remitido giran a velocidades superiores a 100.000 r.p.m., lo que es posible gracias a que el momento de inercia de las peonzas respecto a su eje de figura es muy pequeño y la energía que reciben en el lanzamiento da por resultado una velocidad de rotación muy elevada. Las peonzas más corrientes en forma de pera tienen momentos de inercia mucho mayores, por lo que la energía de lanzamiento provoca velocidades de rotación bastante más reducidas.

La peonza más interesante con que me he topado hasta ahora es la peonza Tippe, que les describí en diciembre de 1979. El trompo en cuestión presenta



Fotografía de señales de hollín sobre una peonza Tippe realizada por Frank F. Johnson



Señal dejada por una de las peonzas Tippe de Johnson sobre vidrio recubierto de hollín

una mitad inferior semiesférica y un vástago central. Para hacerla bailar, se coloca con la semiesfera en la parte inferior, posición lógica dado que es la porción más pesada de la peonza. No obstante, apenas se suelta, la peonza se invierte y se pone a bailar sobre el vástago, elevándose la porción semiesférica en un aparente desafío a la gravedad.

La causa de esta extraña inversión está en el rozamiento con la superficie sobre la que gira la semiesfera. Haciendo bailar la peonza Tippe sobre una hoja de vidrio recubierta de hollín se consigue un dibujo que proporciona una visión de conjunto de su trayectoria, que se reproduce en la ilustración superior de esta página. (Fue obtenida y estudiada en 1960 por Frank F. Johnson, de Hasbrouk Heights, New Jersey.)

¿Cómo se comportaría una peonza sobre una superficie levemente inclinada? ¿Se desplazaría el punto de contacto exactamente igual que sobre una superficie horizontal? Ledo Stefanini, de la Universidad de Bolonia, acaba de publicar una respuesta detallada a estas preguntas. Al igual que otros estudiosos de la influencia del rozamiento sobre el comportamiento de las peonzas, Stefanini ha supuesto que el pico era semiesférico.

Si el trompo gira sin precesión, se desplazará horizontalmente a través del plano inclinado describiendo una recta. Pero si precesiona, hay cuatro posibilidades según la pendiente del plano inclinado y la inclinación de la peonza respecto a la vertical. Una de las trayectorias posibles del pico es una espiral plano abajo. Las otras trayectorias se asemejan a los modelos de nutación de una peonza ideal de pico agudo que no se desplace por el suelo. En algunos casos, el pico traza rizos sobre el suelo a la vez que se desplaza preferentemente en dirección horizontal. También puede dibujar guirnaldas. Para otro grupo de ángulos, el pico empieza trepando plano arriba y luego desciende moviéndose horizontalmente a través del plano.

Quizá resulte de interés ensayar distintas combinaciones de ángulos en un plano inclinado. Stefanini ha publicado un calco que obtuvo con una peonza zigzagueando sobre papel carbón adherido a un plano inclinado. Cualquiera puede hacer sus propios calcos para clasificar los movimientos posibles.

Libros

Mayoría de edad de la ecología, importancia de la farga y riqueza del paisaje vegetal catalán

María Rosa Miracle, Salvador Alegret y Jaume Terradas

CONCEPTOS UNIFICADORES EN ECOLOGIA. W. H. van Dobben y R. H. Lowe-McConnell (editores); Editorial Blume, 397 págs., Barcelona, 1980. No hay duda de que la ciencia ecológica ha experimentado recientemente una verdadera revolución en la manera de plantear sus conocimientos y adquisiciones. La ecología, hasta mediados de siglo, no era más que una historia natural en la que se tendía a enumerar y describir una serie de hechos más o menos anecdóticos, donde se prestaba un interés mayor por las diferencias o discrepancias entre ellos que por las similitudes. La ecología, como ciencia integradora, requirió, para su desarrollo, la formación anterior de otras ciencias biológicas y un conocimiento detallado del medio físico. El disponer de una instrumentación adecuada para llevar el "laboratorio" al campo, como legado de la gran expansión tecnológica que siguió a la segunda guerra mundial, hizo que se recopilaran datos en multitud de situaciones y sistemas naturales. A partir de aquí se pasa, de mera yuxtaposición de descripciones, a una ciencia que busca la formulación de principios generales. Este, en mi opinión, constituye el gran salto en la historia de la ecología, que ha llevado al establecimiento de una ciencia *stricto sensu*. Esa búsqueda de regularidades, de conceptos o leyes generales, de propiedades o fenómenos que se repiten en todos los ecosistemas, ha constituido el cambio necesario para que se pueda empezar a hablar de una ciencia ecológica. Prueba de ello es que hasta 1974 no se organizó el primer congreso internacional de ecología, cuando ya se había alcanzado el nivel de abstracción necesario, que se había estado gestando en la década anterior.

El título del libro, *Conceptos unificadores en ecología,* refleja el tema principal de ese primer congreso internacional, cuyas sesiones plenarias estuvieron dedicadas precisamente a los distintos logros que posibilitaron el grado de abstracción requerido por toda ciencia, que permitían una unificación o generalización de los conceptos. Se recogen las comunicaciones presentadas en las sesiones plenarias del congreso, celebrado en La Haya, en septiembre de 1974, organizado por la International Association of Ecology (INTECOL) bajo los auspicios de la International Union of Biological Sciences (IUBS). Durante una semana reunió a ochocientos ecólogos de muchos países, invitándose a las figuras más conspicuas en el campo de la ecología a que presentaran sendas comunicaciones en unas sesiones plenarias. Otras comunicaciones no solicitadas, que también se leyeron o presentaron, han sido objeto de un volumen aparte: "Proceedings of the First International Congress of Ecology: structure, functioning and management of ecosystems".

Conceptos unificadores en ecología compendia un total de veinte comunicaciones (además de algunos comentarios cortos) con nuevas aportaciones teóricas obtenidas, en general, de la integración de la labor realizada a lo largo de varios años de investigación. Los autores, en su mayoría, son los pioneros o creadores de la teoría ecológica: Holling, Margalef, May, Odum, Whittaker, entre otros. Publicado en español, como volumen 11º de la acertada serie de títulos de "Blume ecología", el libro responde a la filosofía de dicha colección: poner a disposición del público una información seria y profunda sobre temas ecológicos.

Las comunicaciones se distribuyen en cinco temas principales, correspondientes a las sesiones del congreso, incluyendo, al final de cada uno de ellos, un resumen de las discusiones suscitadas. El primer tema, "Flujo de energía y materia entre niveles tróficos", comprende cinco artículos, con puntos de vista bastante diferentes. Unos lo enfocan teóricamente, considerando sistemas ecológicos completos, otros presentan problemas y resultados concretos, como el papel de los descomponedores y el de los consumidores. Se discute la validez de ciertos conceptos, el de nivel trófico, por ejemplo, pues cuanto más alejado se halla un organismo de la entrada de energía en el ecosistema más pequeña es la probabilidad de que obtenga su alimento de un único nivel trófico. Sin embargo, la mayoría coincide en considerar una organización jerárquica de los componentes tróficos en la que los ecosistemas consiguen la estabilidad y la persistencia necesarias.

"Productividad comparada en los ecosistemas" es el título del segundo tema, que agrupa cinco escritos dedicados principalmente a evaluar las estimas existentes de las producciones primaria y secundaria en la biosfera y presentar modelos que permitan calcular o predecir la producción que potencialmente puede proporcionar nuestro planeta. Todos los trabajos se refieren a la producción en ecosistemas, excepto el último, que trata del problema del control o gestión de las proliferaciones explosivas de algas (sobre todo cianofíceas) en ecosistemas acuáticos eutróficos.

El tercer tema, "Diversidad, estabilidad y madurez en los ecosistemas naturales", compendia cuatro comunicaciones que tratan de definir, medir e interrelacionar las citadas características de los ecosistemas. Margalef, en una de las comunicaciones de esta sección, aporta interesantes generalizaciones sobre la tendencia de la naturaleza a hacer, a lo largo del tiempo, las relaciones cada vez más complejas y predictibles, y, más lenta, la tasa de renovación, lo que significa un gasto menor de energía por biomasa mantenida. Asimismo ilustra cómo dichas propiedades (diversidad, estabilidad y madurez) se corresponden en razón de su dependencia común de las leyes de la termodinámica y de la selección natural. Esta misma idea sobre qué regulan los sistemas vivos está presente en otros escritos del autor, hasta dar título a su última obra: *La Biosfera: entre la termodinámica y el juego*.

Las discusiones en torno al papel del hombre en la biosfera y la gestión de los ecosistemas constituyeron una parte esencial del congreso, cuyas dos últimas sesiones estuvieron dedicadas a estos temas. Bajo el título "Diversidad, estabilidad y madurez en ecosistemas influidos por las actividades humanas" se incluyen dos artículos y dos escritos más breves, a modo de comentarios. Uno de ellos viene a ser una compilación de pruebas o ejemplos de cómo el hombre interfiere en las propiedades de los ecosistemas en todos los sentidos. El otro aporta un modelo teórico que intenta predecir el efecto de las perturbaciones sobre la estabilidad de los ecosistemas. Finalmente, en la últi-

ma sección, "Estrategias para la gestión de los ecosistemas naturales y artificiales", se pueden encontrar algunas descripciones de casos concretos de gestión en países tropicales y europeos, y notas y comentarios sobre los tipos de técnicas de gestión; en lo concerniente a estas últimas se examinan las pretéritas y las presentes, al tiempo que se insiste en la creación de una ciencia de gestión ecológica y se avanza un programa de la misma. Es evidente que todavía estamos lejos de poder dar reglas exactas para cada problema de gestión y que la utilización de recursos vulnerables exige una planificación global a largo plazo. (M. R. M.)

LA FARGA, por Pere Molera i Solà. Editorial Dopesa; 111 págs. 35 ilustraciones, Barcelona, 1980. Aseguraría que la mayoría de lectores españoles cuando se adentran en obras de tanto prestigio como la *Encyclopedia of Chemical Technology,* de Kirk-Othmer, *The Making, Shaping and Treating of Steel,* de United States Steel, o el *Nouveau Traité de Chimie Minérale,* redactado bajo la dirección de Paul Pascal, y hallan citado, de forma destacada, el *catalan process,* el *catalan furnace* o el *bas foyer catalan* como uno de los procedimientos directos de obtención de hierro, se ven sorprendidos escépticamente. ¿Realmente fue la farga catalana un proceso metalúrgico tan importante? Pere Molera, profesor de metalurgia en la Universidad Autónoma de Barcelona, responde a esta pregunta en la obra que aquí reseñamos. El autor nos introduce en la metalurgia autóctona del hierro, no sólo desde una perspectiva historiográfica, sino también, y de modo particular, desde un punto de vista químico y metalográfico, dándonos a conocer, con un lenguaje ameno, los resultados de sus ensayos sobre los distintos materiales que salían de las fargas. Estudios efectuados sobre rejas, llaves, clavos, armas, utensilios, etc., muchos de los cuales han permanecido durante siglos en la intemperie, han revelado que la calidad de los productos de las fargas era superior a la mayoría de las aleaciones férreas actuales. Por tanto, el estudio de la farga desde una óptica actual no es un ejercicio sentimental, sino que queda justificado por su interés como fundamento de un proceso metalúrgico alternativo.

El procedimiento catalán era del tipo de reducción directa a temperatura inferior a la de la fusión; desde un punto de vista energético, necesitaba menos combustible que los procedimientos actuales que conllevan la fusión de la masa. La farga consistía esquemáticamente en mezclar de forma conveniente la mena de hierro pulverizada con carbón vegetal en una especie de horno, donde se introducía una tobera; después de la reducción, se martillaba el producto para eliminar las escorias y aumentar su compacticidad. El horno, que constituía el elemento más importante, poseía forma y dimensiones precisas. El gran secreto residía en la longitud de la tobera que introducía aire en el fondo del horno. La trompa, un ingenioso procedimiento hidráulico de inyectar el aire, era el elemento más espectacular, aunque de escasa relevancia tecnológica.

La calidad del producto manufacturado fue la clave de la expansión del procedimiento catalán, tanto en Europa como en América, hasta finales del siglo XIX. La aparición de los altos hornos paralizó rápidamente las fargas, como consecuencia de la mayor rentabilidad de la innovación tecnológica, pero a costa de una pérdida de calidad, no recuperada hasta el día de la fecha.

El autor capta totalmente la atención del lector cuando enumera los puntos de contacto entre la farga y la siderurgia moderna. En efecto, el lingote de alto horno, como es sabido, es descarburado y purificado en los hornos eléctricos de las acerías. No obstante, la cantidad creciente de chatarra y la disponibilidad de materiales, como la esponja de hierro o prerreducidos, junto con el encarecimiento del coque, van aconsejando la utilización de dichos hornos en la fabricación de hierro en estado sólido o pastoso; procedimiento fundamentalmente parecido al de la farga. La técnica pulvimetalúrgica de la forja de preformas sintetizadas guarda gran similitud con la farga; en ambas técnicas se trata de un proceso de densificación plástica del metal caliente.

Señalemos, por último, que el trabajo de Molera viene a llenar el vacío existente sobre el tema en la bibliografía científica de expresión latina. Esperemos que a partir de ahora la farga catalana sea más conocida dentro de nuestras fronteras y que publicaciones de nuestra área cultural no confundan más la farga (milenario proceso de obtención del hierro) con la forja (procedimiento de conformación metálica por compresión en caliente). (S. A.)

LA VEGETACIÓ DELS PAÏSOS CATALANS, por R. Folch. Editorial Ketres; Barcelona, 1981. El investigador que más ha contribuido al conocimiento científico del paisaje vegetal catalán, O. de Bolòs, afirma en el prólogo que había llegado ya el momento de realizar una obra como ésta. Ello es, sin duda, cierto, y aún me atrevería a decir que dicho momento llegó hace algunos años, aunque quizá por entonces no hubiese sido posible un resultado tan completo como el que ahora tenemos. Son bastantes las promociones de jóvenes naturalistas que han sentido esta carencia y a los que no les bastaba la existencia de algunas obras fitocenológicas de ámbito comarcal para hacerles accesible el tema. Entre los breves resúmenes, como los excelentes del propio Bolòs en *Geografia de Catalunya* (Ed. Aedos) o *Geografia física dels Països Catalans* (Ketres), y las monografías comarcales o los artículos especializados existía un salto muy difícil de dar.

Por supuesto, una de las mayores dificultades estriba en que la descripción de la vegetación se basa en inventarios florísticos: las unidades del paisaje se describen y definen a partir de la composición específica. Quien no sea un consumado botánico poco puede deducir de largas listas de nombres científicos de especies. Para acabar de complicar las cosas, sólo se dispone todavía de floras áridas e incompletas por estar centradas en países vecinos pero distintos. No es, pues, de extrañar que la situación del principiante haya sido durante años descorazonadora.

Creo que es ésta una buena ocasión para un comentario de orden general. Muchos investigadores son, por temperamento y por características de su trabajo, gentes que se encierran en problemas poco comprensibles o poco atractivos para la mayoría de la población. Convencidos de la importancia de su tarea, creen con razón que los frutos que obtengan se extenderán en el futuro de un modo u otro al colectivo social; consideran prioritario por tanto avanzar en el conocimiento científico fundamental, y en ello ponen un empeño y una dedicación muy meritorios. En nuestro país no han faltado, por suerte, hombres de este tipo a cuyos nombres podemos asociar el nivel de conocimiento alcanzado en los diversos campos de la ciencia. Pero no cabe duda de que algo funciona mal en nuestra vida cultural. Las individualidades suelen quedar en eso, singularidades cuyo esfuerzo se proyecta demasiado tenuemente en el ámbito social. Las causas son complejas, y no podemos analizarlas aquí, pero señalemos que la propia universidad parece siempre dispuesta a postergar, con demasiada facilidad, las tareas de extensión que en teoría debiera llevar a cabo, y que muchos universitarios tienden a considerar los manuales más como una especie de negocio personal de algunos que como una labor creativa y útil.

Todo ello viene a cuento para explicar la valoración que hacemos del libro de Folch en el contexto del estudio científico del paisaje vegetal. Nuestros botánicos, por las razones expuestas al principio, se han encontrado siempre aislados como si de sacerdotes de un culto más o menos vestigial se tratara. Para tratar de salir de tal situación, y aunque el conocimiento científico de nuestra flora y vegetación les requiere todavía un esfuerzo titánico, o tal vez comprendiendo que dicha tarea escapa a las posibilidades de unas pocas personas, han hecho reiterados intentos de comunicación con un público más amplio, intentos que a mi juicio son buena prueba de la sensibilidad de estos investigadores, bastante excepcional en el contexto en que nos movemos. Entre los trabajos más relevantes citemos dos libros, *El paisaje vegetal barcelonés,* de O. de Bolòs, y *L'alta muntanya catalana: flora i vegetació,* de J. Vigo. En este último hallamos ya un afortunado esfuerzo por dar a la introducción al estudio del paisaje vegetal la necesaria base florística elemental. El libro de Folch, un poco en la línea de este último pero cubriendo un ámbito territorial mayor, tiene también la virtud de ofrecer, además de la información fitocenológica, los elementos necesarios para una identificación suficiente desde el punto de vista florístico y estructural de las unidades paisajísticas. Aunque sólo fuera por su carácter de síntesis, y no por las aportaciones personales abundantes de su autor, el libro habría de convertirse en punto de partida y referencia obligada para el futuro. Al facilitar el buen conocimiento del paisaje, será pieza útil e importante de desarrollo obligada para el futuro.

Así pues, desde el punto de vista de la realidad cultural en que nos movemos, Folch ha dado de nuevo en la diana. Recordemos el *Llibre Blanc de la Natura,* del que si no fue autor único, sí fue sin duda acertado promotor y coordinador, como de otras obras menores presididas por parecido espíritu y de un sinnúmero de actividades que le han convertido en nuestro naturalista con mayor audiencia en la administración catalana, para resaltar que esta obra constituye ahora un hito muy destacado en el campo específico del conocimiento del paisaje.

Junto a Folch, el libro tiene otros dos autores, dibujantes y botánicos a la vez, como era indispensable. E. Sierra, ilustrador también con igual maestría del ya citado libro de Vigo, y J. Nuet, cuyos esquemas estructurales sobre todo constituyen una novedad del mayor interés. La calidad y rigor de la ilustración contribuyen muchísimo a hacer el libro excepcional en la literatura fitocenológica, incluso a nivel internacional, pero no debemos entenderlo como un lujo, sino como una herramienta para recuperar el retraso que padecemos en el conocimiento de nuestra propia naturaleza.

La estructura del libro es relativamente sencilla. Tras algunos capítulos de introducción a la fitocenología y de diversos conceptos y términos biogeográficos, que ocupan sólo las primeras cincuenta páginas, la práctica totalidad del texto se dedica a la flora y comunidades vegetales. Se parte del reconocimiento de cinco grandes unidades territoriales (tierra baja mediterránea; línea litoral; montaña mediterránea y submediterránea culminal; montaña media lluviosa submediterránea, centroeuropea y atlántica; y alta montaña subalpina y boreoalpina). Para cada una, tras ciertas consideraciones generales, se repasan los dominios climácicos y los ambientes particulares (vegetación de ribera y de las aguas, rupícola, arvense, nitrófila, etc.). En todos los casos se describen las comunidades, con mención de las especies que las componen habitualmente, descripción de la estructura global de la comunidad y de la distribución geográfica en el territorio considerado. Todo ello se ilustra abundantemente con dibujos de las especies, bloques estructurales, mapas de distribución, perfiles y fotografías. Los títulos de los apartados corresponden a nombres de comunidades en catalán y, cuando es posible, llevan la denominación científica entre paréntesis.

El libro contiene como apéndices una tabla de comunidades, que recoge la sistematización fundamentalmente elaborada por O. de Bolòs y J. Vigo, y un índice de referencias bibliográficas de estas comunidades (un total superior a 500 asociaciones) con indicación de los trabajos en que fueron descritas y de las aportaciones posteriores hasta fines de 1979, índice que, por sí solo, supondría ya una contribución muy útil para los estudiosos del paisaje. Hay también un nomenclátor fitocenológico que pone particularmente de relieve uno de los valores destacados de la obra: su aportación en materia de normalización lingüística. Destaquemos también que la editorial ha hecho un esfuerzo considerable, cuidando mucho presentación y texto.

Una crítica no debe ser sólo laudatoria, ni siquiera cuando se trata de una obra necesaria y tan satisfactoriamente resuelta. Señalemos que los capítulos introductorios se resienten de un planteamiento demasiado convencional, particularmente en lo relativo a la sucesión y a conceptos como el de clímax. La propia opción por un tratamiento estrictamente limitado a la fitosociología podría discutirse; aunque no cabe duda que se beneficia así de casi todo el esfuerzo científico realizado hasta ahora entre nosotros en el estudio del paisaje, resulta a veces, para mi gusto, excesivamente sistematizador y poco explicativo. Tal vez envidiamos un poco la habilidad que muestran no pocos textos, particularmente ingleses, para resaltar lo esencial incorporando explicaciones funcionales y desechando un detalle puntilloso en la enumeración de comunidades. En un nivel medio, como el de este texto, quizá se hubiera podido renunciar a las ambiciones de catálogo exhaustivo. Son, sin duda, estas ambiciones de catálogo las que han conducido a un tratamiento demasiado superficial de cada comunidad, que sabe a poco en las más importantes y es prácticamente reducible al nombre en algunas de las menores. Ello puede resultar decepcionante cuando se busca información relativa a partes concretas del territorio considerado. La introducción de un enfoque más ecológico, con la ayuda de los múltiples trabajos realizados, ciertamente en otros países, pero sobre comunidades funcionalmente muy similares, hubiese dado mayor atractivo al texto.

El autor, buen conocedor del idioma, tiende a veces a un cierto barroquismo literario que puede hacer perder claridad. A ello hay que unir que ciertos capítulos, como el que trata de los elementos corológicos, resultan demasiado breves y en exceso cargados en terminología para el público al que se dirigen. No obstante, el conjunto es muy inteligible, y ya nos hemos referido al esmero con que se viene a contribuir en los aspectos lingüísticos.

Personalmente, hubiese agradecido hallar el índice de materias al principio y no al final del texto; creo que ello hubiese facilitado su inevitablemente frecuente consulta, dada la gran cantidad de unidades de paisaje descritas. Pero también esto es opinable y, como lo anterior, no debe tomarse en detrimento del conjunto. En realidad, el más grave reparo es de otro tipo. El precio del libro es ajustado al esfuerzo realizado y perfectamente justificable, pero lo cierto es que resulta excesivo para una parte muy importante de sus destinatarios. A pesar de lo cual, debemos desde luego congratularnos de la aparición de una obra que habrá de representar un importante papel formativo y que es, ya desde ahora, indispensable. (J. T.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

ENERGIA SOLAR FOTOVOLTAICA

SOLAR CELLS. C. E. Backus. IEEE Press Book, 1976.

SOLAR ENERGY CONVERSION: THE SOLAR CELL. R. C. Neville. Elsevier, 1978.

SOLAR ELECTRICITY, AN ECONOMIC APPROACH TO SOLAR ENERGY. W. Palz. UNESCO, 1978.

THIRTEEN PHOTOVOLTAIC SPECIALIST CONFERENCE. IEEE Press; Washington, 1978.

SECOND EUROPEAN COMMUNITY PHOTOVOLTAIC CONFERENCE; Berlín, 1979.

EL ARMAZON CELULAR

THREE DIMENSIONAL FINE STRUCTURE OF CULTURED CELLS: POSSIBLE IMPLICATIONS FOR SUBCELLULAR MOTILITY. Ian K. Buckley en *Tissue and Cell,* vol. 7, n.° 1, págs. 51-72; 1975.

MICROTRABECULAR LATTICE OF THE CYTOPLASMIC GROUND SUBSTANCE: ARTIFACT OR REALITY. John J. Wolosewick y Keith R. Porter en *The Journal of Cell Biology,* vol. 82, n.° 1, págs. 114-139; julio, 1979.

THE CONTROL OF PIGMENT MIGRATION IN ISOLATED ERYTHROPHORES OF *HOLOCENTRUS ASCENSIONIS* (OSBECK), I: ENERGY REQUIREMENTS. Katherine J. Luby y Keith R. Porter en *Cell,* vol. 21, n.° 1, págs. 13-23; agosto, 1980.

LAS ERUPCIONES DEL MONTE STA. ELENA

VOLCANOLOGY. Howel Williams y A. R. McBirney. Freeman, Cooper and Co., 1979.

FIRE AND ICE: THE CASCADE VOLCANOES. Stephen L. Harris. Pacific Search Press, 1980.

1980 ERUPTIONS OF MOUNT ST. HELENS. Dirigido por Peter W. Lipman y Donal R. Mullineaux. U.S. Geological Survey Professional Paper n.° 1250, 1981.

LA GALAXIA DE LA VIA LACTEA

THE LARGE-SCALE CHARACTERISTICS OF THE GALAXY. Dirigido por W. B. Burton. D. Reidel Publishing Company, 1979.

INTERSTELLAR MOLECULES. Dirigido por B. H. Andrew. D. Reidel Publishing Company, 1980.

STAR CLUSTERS. Dirigido por James E. Hesser. D. Reidel Publishing Company, 1980.P

THE MILKY WAY. Bart J. Bok y Priscilla F. Bok. Harvard University Press, 1981.

POLINIZACION SUBMARINA

BIOLOGY OF AUSTRALIAN SEAGRASSES: POLLEN DEVELOPMENT AND SUBMARINE POLLINATION IN *AMPHIBOLIS ANTARCTICA AND THALASSODENDRON CILIATUM* (CYMODOCEACEAE). Sophie Ducker, John Pettitt y Bruce Knox en *Australian Journal of Botany,* vol. 26, págs. 265-268; 1978.

REPRODUCTION IN SEAGRASSES: NATURE OF THE POLLEN AND RECEPTIVE SURFACE OF THE STIGMA IN THE HYDROCHARITACEAE. John Pettitt en *Annals of Botany,* vol. 45, págs. 257-271; 1980.

PERCEPCION ANORTOSCOPICA

POST-RETINAL VISUAL STORAGE. Theodore E. Parks en *The American Journal of Psychology,* vol. 78, n.° 1, págs. 145-147; marzo, 1965.

DISTORTIONS IN MOVING FIGURES VIEWED THROUGH A STATIONARY SLIT. S. M. Anstis y Janette Atkinson en *The American Journal of Psychology,* vol. 80, n.° 4, págs. 572-585; diciembre, 1967.

INTELLIGENCE FACTORS IN THE PERCEPTION OF FORM THROUGH A MOVING SLIT. I. Rock y E. Sigman en *Perception,* vol. 2, n.° 3, págs. 357-369; 1973.

AN INTRODUCTION TO PERCEPTION. Irvin Rock. Macmillan Publishing Co., 1975.

BIOQUIMICA DE LA RESISTENCIA A LA MALARIA

MALARIA AND HUMAN POLYMORPHISMS. Frank B. Livingstone en *Annual Review of Genetics,* vol. 5, págs. 33-64; 1971.

CULTIVATON OF MALARIAL PARASITES. William Trager y James B. Jensen en *Nature,* vol. 273, n.° 5664, págs. 621-622; 22 de junio de 1978.

GENETICS OF RED CELLS AND SUSCEPTIBILITY TO MALARIA. L. Luzzatto en *Blood,* vol. 54, n.° 5, págs. 961-976; octubre, 1979.

EL DESCUBRIMIENTO NEWTONIANO DE LA GRAVEDAD

FROM KEPLER'S LAWS, SO-CALLED, TO UNIVERSAL GRAVITATIONS: EMPIRICAL FACTORS. Curtis A. Wilson en *Archive for History of Exact Sciences,* vol. 6, n.° 2, págs. 89-170; 29 de abril de 1970.

FORCE IN NEWTON'S PHYSICS: THE SCIENCE OF DYNAMICS IN THE SEVENTEENTH CENTURY. Richard S. Westfall. American Elsevier Publishers Inc., 1971.

THE NEWTONIAN REVOLUTION: WITH ILLUSTRATIONS OF THE TRANSFORMATION OF SCIENTIFIC IDEAS. I. Bernard Cohen. Cambridge University Press, 1980.

TEMAS METAMAGICOS

GROUPS AND THEIR GRAPHS. I. Grossman y W. Magnus. New Mathematical Library, 1975.

TALLER Y LABORATORIO

THE TOP: UNIVERSAL TOY, ENDURING PASTIME. D. W. Gould. Clarkson N. Potter, Inc., 1973.

A REALISTIC SOLUTION OF THE SYMMETRIC TOP PROBLEM. T. R. Kane y D. A. Levinson en *Journal of Applied Mechanics,* vol. 45, n.° 4, págs. 903-909; diciembre, 1978.

BEHAVIOR OF A REAL TOP. Ledo Stefanini en *American Journal of Physics,* vol. 47, n.° 4, págs. 346-350; 1979.

ERRATA

En el número 55, correspondiente al mes de abril de este año, en la página 17, última línea de la tercera columna, donde dice "unos 6025 milímetros de agua" debe decir "unos 2440 milímetros de agua"; en la página 19, primera línea de la primera columna, donde dice "Unos 2260 suplen" debe decir "Unos 915 suplen"; en la página 20, sexta línea del último párrafo, donde dice "esporas reproductoras" debe decir "células reproductoras".